DEVELOPING A NATIONAL ASSESSMENT MODEL TO INFORM EDUCATIONAL POLICY IN BHUTAN

Submitted by

Gembo Tshering

B.Ed, PGCTIS, M.Sc

A thesis submitted in total fulfillment
of the requirements for the degree of
Doctor of Philosophy

School of Education

Faculty of Education

La Trobe University

Bendigo, Victoria, 3552

Australia

August 2012

Goley Dukpa, Meme Nima, and Ninda Lhamo

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND TERMS

ACER	Australian Council for Educational Research
AERA	American Educational Research Association
ANXMAT	Index of Anxiety
APA	American Psychological Association
ATSCHL	Index of Attitude Towards School
BBE	Bhutan Board of Examinations
B.Ed	Bachelor of Education
BELONG	Index of Belongingness to School
BIB	Balanced Incomplete Block
BPC	Board of Participating Countries
CAPSD	Curriculum and Professional Support Service Division
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CI	Confidence Interval
ClL	Classroom Level
CITO	National Institute of Educational Measurement (Dutch)
CoL	Context Level
COMPHOME	Computer facilities at home
COMPLRN	Index of Competitive Learning Preference
COOPLRN	Index of Cooperative Learning Preference
CSTRAT	Index of Control Strategies
CTPSO	Index of Constructivist Teaching Approach
CTPEL	Index of Extended Teaching Approach
CTPST	Index of Structured Teaching Approach
CULTPOSS	Cultural possession
DIF	Differentially Functioning Items
DISCLIM	Index of Classroom Disciplinary Climate
ELAB	Index of Elaboration
EMIS	Educational Management and Information System
ETS	Educational Testing Service
FWEALTH	Family wealth
FPC	Framework Planning Committee
FSC	Framework Steering Committee
GNH	Gross National Happiness
HERDES	Home educational resources
HIGHCONF	Confidence in Performing High-Level Tasks
HISEI	Highest Occupational Status of One of the Parents
HOMPOSS	Home Possession
ICT	Information and Communications Technology
IEA	International Association for the Evaluation of Educational Achievement
InL	Input Level
INMAT	Index of Extrinsic Motivation
INSMOT	Index of Intrinsic Motivation

INTCONF	Index of Confidence in Performing Internet Tasks
INTUSE	Index of Internet and Entertainment
IRT	Item Response Theory
ISCE	International Standard Classification of Education
ISCO	International Standard Classification of Occupations
ISEI	International Socio-Economic Index of occupational status
LEP	Limited English Proficiency
LISREL	Linear Structural Relations
M.Sc	Master of Science
MATHEFF	Index of Self-Efficacy
MCAR	Missing Completely At Random
MEMOR	Index of Memorization
MLE	Maximum Likelihood Estimation
MNSQ	Mean Square Fit Statistics
MVA	Missing Value Analysis
NAEP	National Assessment of Educational Progress
NAGB	National Assessment Governing Board
NCME	National Council on Measurement in Evaluation
NCTM	National Council of Teachers of Mathematics
NEA	National Education Assessment
NIER	National Institute for Educational Policy Research
NNFI	Non-Normed Fit Index
NUFFIC	Netherlands Organization for International Cooperation in Higher Education
OECD	Organization for Economic Co-operation and Development
OuL	Output Level
PARED	Higher Parental Education Level in terms of Years of Schooling
PGCE	Post-Graduate Certificate in Education
PGCTIS	Post-Graduate in Teaching of Information Systems
PIRMS	Principal Management Instructional Scale Rating
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PRGUSE	Index of Computer Programme and Software
RMR	Root Mean Square Residual
RMSEA	Root Mean Square Error of Approximation
ROUTCONF	Index of Confidence in Performing Routine Tasks
ScL	School Level
SCMAT	Index of Self-Concept
SES	Socio-Economic Status
SMB	School Management Board
SPSS	Statistical Package for the Social Sciences
StL	Student Level
STUDBEHA	Index of Student Behavioural Factors
STUREL	Index of Student and Teacher Relationship
TALIS	Teaching and Learning International Survey

TCCONS	Index of Teacher Consensus
TCDC	Index of Classroom Climate
TCMORALE	Index of Teacher Morale and Commitment
TCONRT	Index of Resource-Related Constraints
TCONST	Index of Student-Related Constraints
TDCT	Index of Beliefs about Constructivist Teaching Approach
TDTM	Index of Beliefs about Direct Transmission Teaching Approach
TEACHBEHA	Index of Teacher Behavioural Factors
TEACHSUP	Index of Teacher Support
TIMSS	Trends in International Mathematics and Science Study
TPC	Index of Teacher Professional Collaboration
TREM	Index of Re-enforcements
TSCHL	Index of School Climate
TSE	Index of Teachers' Self-efficacy
U.K	United Kingdom
UNESCO	United Nations, Scientific and Cultural Organization
U.S.	United States
WLS	Weighted Least Squares

LIST OF APPENDICES (Refer to CD in pocket inside back cover)

SUMMARY

Student, teacher, school, and environmental characteristics constitute the fundamental building blocks of a nation’s education system. Knowledge about these characteristics can help stakeholders in a nation’s education system to improve its quality. However, understanding how these characteristics impact upon an education system remains a major challenge in many countries.

This study developed a national educational assessment model capable of generating knowledge about student, teacher, school, and contextual characteristics of the Bhutanese education system. The study used a cross-sectional survey and a focus group interview to collect data. One thousand five hundred students, 60 teachers, and 60 school principals, selected by using a two-stage cluster sampling method, participated in the survey. Three people, based on a purposeful sampling method, from the Ministry of Education of the Royal Government of Bhutan participated in the focus group interview.

Results from the study indicated that students’ demographic profile, motivation, self-beliefs, self-regulation, and learning preferences, as well as their experience with ICT, homework, classroom management, and school climate were related to their achievement. Similarly, teachers’ demographic profile, professional development, appraisal and feedback, and self-efficacy correlated with student achievement. Further, teachers’ experience with school climate, classroom management, teaching, homework, and tests also related to student achievement. However, school policies and practices, climate, resources, and educational leadership had a modest bearing on student achievement. Finally, the analyses of contextual data showed that Bhutan has a great opportunity for stimulating the performance of its education system further by considering and implementing more national-level, research-based interventions.

Overall, an emerging theme from the study indicates that teaching and learning are most effective when student, teacher, school, and contextual characteristics in a nation’s education system complement one another. This theme underscores the importance of considering these characteristics when developing and sustaining national educational policies and interventions.

STATEMENT OF AUTHORSHIP

Except where reference is made in the text of the thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis submitted for the award of any other degree or diploma.

No other person's work has been used without due acknowledgement in the main text of the thesis.

This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution.

DECLARATION OF ETHICS COMMITTEE APPROVAL

All research procedures reported in the thesis were approved by the La Trobe University, Bendigo, Human Ethics Committee (Approval Number FHEC No. R038/09 dated 15 October 2009).

Signed: Dated:

ACKNOWLEDGEMENTS

As learning and its products are always a result of the exchange of information between various parties, so too is this thesis.

My former teachers, lecturers, and professors deserve a large thank-you for giving me the experience of their high expectation, their care for my development, and their willing support that continues to nurture my love of life-long learning.

The Ministry of Education of the Royal Government of Bhutan created the enabling conditions for preparation of this thesis by granting me four years of study leave, giving me access to their records, and allowing me to conduct field research in their schools. I am indebted to them for this, and to the erstwhile Bhutan Board of Examinations for the opportunities it provided in charging me with duties and responsibilities that provided perspectives for this thesis.

My thanks are due to the OECD for allowing me to use test and questionnaire items from its PISA and TALIS programmes. Likewise, thanks to Professor Philip Hallinger, Mahidol University, Thailand for permitting me to use his PIMRS in this study.

In Bhutan, students, teachers, principals and officials gave enthusiastic co-operation and expended much time on providing a rich data source for this thesis. I thank them for this wonderful contribution.

In the Netherlands, Dr Barnard Velkamp at the University of Twente and Dr Bas Hemker at CITO made valuable comments on Chapter 5. In addition, Dr Hans Vox at the University of Twente made valuable suggestions on statistical procedures. In Australia, Dr Doug McCurry at ACER was of great assistance in initiating me into the ConQuest program. Also, Graeme Withers, a former Senior Research Fellow at ACER, made many editorial comments and content feedback that made this thesis much more readable. I thank them for their valuable contributions.

La Trobe University accepted my PhD candidature and offered two extremely valuable scholarships: the La Trobe University Post-Graduate Research Scholarship and the La Trobe Endeavour International Postgraduate Research Scholarship. These brought peace of mind, and instilled purpose into my work, turning my life-long dream of pursuing a PhD into a reality. I will always be indebted to La Trobe University for this. The staff of the School of Education at the University's Bendigo campus provided academic and interpersonal support, and I have felt very much an accepted member of the School of Education "family".

Professor Vaughan Prain and Dr Peter Cox, **my research supervisors**, in particular have, on an almost daily basis, provided support, guidance, inspiration, and wise mentorship to help me mould the thesis into its final form. I have learned so much from them—the instruction was valuable, the relations were easy and cordial, and the deep and vast knowledge I could tap into was of inestimable value. Between them, they matched my constant desire for a perfect Guru.

Last but not least, I thank my family (Yeshey, Pema, Jigme, and Kuenza) for their constant understanding whenever I became preoccupied with the studies. Yeshey made our stay in Australia much more enjoyable by supplementing my scholarship stipend with her hard-earned wages from part-time jobs. Pema has shown a wonderful ability to self-manage his own studies, to love learning, and to communicate with his Australian mates in an Australian school. Away from us, Jigme has always topped her class and Kuenza has begun her schooling with lots of curiosity—the news that always reminded us of the kind and caring support provided to them by our loving relatives, when being away from us. The Royal Government of Bhutan looks to all its citizens to improve the now-famous concept of “Gross National Happiness” of the nation—Pema, Jigme, and Kuenza show every sign of being willing contributors already.

Chapter 1

INTRODUCTION

1.1 Introduction

This study seeks to develop a national educational assessment model for the Kingdom of Bhutan. Such a model would enable the collection and analyses of data on students' knowledge and skills of curricular content and various contextual factors related to effective teaching and learning. Because the success of an education system depends on the quality of its assessment system, a national educational assessment model, of the kind and scope that this study intends to develop, has the potential to function as a barometer of the Bhutanese education system's quality. As presented in later chapters, mathematics will be used as the central criterion subject in this study. The model, however, will be adaptable to other school subjects. In short, this thesis aims to design, conduct, analyse, and report on a national educational assessment programme for Bhutan, giving prompts for research-based policy-making and improved practice.

This chapter commences by looking into the needs for a national educational assessment model in Bhutan, followed by presentation of the research aim that foreshadows and underlies the potential outcomes of the system-wide educational assessment model. The chapter then positions the research aim and outcomes in the context of the Bhutanese education system. This is followed by an overview of the history and the highlights of the current goals and policy directives of the Bhutanese education system. A model of education is then identified that best matches the salient aspects of the Bhutanese education system. The chapter then sets the scene by discussing the current educational assessment programmes in Bhutan and in other countries. The remaining sections of this chapter will present the background to the study, the research methods in the study, the significance and possible limitations of the study, and the thesis overview.

1.2 The Need for a National Educational Assessment Model in Bhutan

The general need for a national educational assessment model in Bhutan arises from the following five broad perspectives: the development of school autonomy; the value for future policy-making and practice of being able to match international benchmarks; the focus on enabling and developing a standards-based school curriculum; the development of relevant research-based evidence; and the need for an assessment database.

First, the recent policy change of the Ministry of Education of the Royal Government of Bhutan of granting more autonomy to school districts allows schools to function differently, which may result in different student outcomes. This means that schools can learn from each other, and parents can more wisely choose schools for their children. For the schools to learn from each other's strengths and weaknesses and for the parents to choose the right schools for their children, a mechanism that presents accurate data on the schools' performances is needed. A national educational assessment model will be able to provide information about, and the factors underlying the differences in, student outcomes.

Second, the proliferation of technology-driven learning opportunities and an increasing global interdependence necessitate that Bhutan be able to benchmark its education system with the education systems of other countries. A national educational assessment model will be able to link and compare the performance of Bhutanese students with the performance of the students from other countries by using and sharing common data collection models.

Third, recently Bhutan started reforming its school curricula by introducing standards-based curricula in its schools. A system-wide educational assessment model will be able to generate profiles of students' knowledge against individual curriculum standards.

Fourth, with the rapid increase in the number of schools in Bhutan, the need to base educational policy decisions on research-based evidence is pressing. A national educational assessment model will provide stakeholders in the Bhutanese education system with wide-ranging information about its health.

Fifth, interest in research in education has been growing steadily in Bhutan in recent years, especially since the establishment of the Royal University of Bhutan in 2003, which is also the maiden university in Bhutan. A database will emerge from the national educational assessment programme, with the data collected for this thesis as an important resource for longitudinal studies and future research on the Bhutanese education system.

In summary, a national educational assessment model will have the capacity to contribute to teachers teaching more effectively, students learning with greater understanding, parents making more informed choices of schools, schools functioning more effectively, and the Royal Government of Bhutan making better informed policy decisions on school education. The proposed model should also bridge the gap between policy directives and field realities, and guide the Bhutanese education system in producing higher quality student outcomes and robust human capital with the capacity and desire for life-long learning.

1.3 Research Aim

This thesis aims to develop a national educational assessment model for Bhutan with an emphasis on its capacity to generate the following 10 outcomes. The outcomes are also the key indicators of the prospective success of this thesis and the bases for analyses of the current performance of the Bhutanese Education System.

Outcome 1: Knowledge profiles of students about their comprehension of school curricular content

Outcome 2: Knowledge about the factors related to effective schooling and their effects on student outcomes

Outcome 3: Knowledge about student outcomes at school, district, regional, and national levels

Outcome 4: Knowledge about the preparedness of Bhutanese students to meet the challenges of the future

Outcome 5: Knowledge about the skills that Bhutanese students need to adapt to rapid societal and technological change

Outcome 6: Knowledge about teachers and teaching and their effects on student outcomes

Outcome 7: Knowledge about the educational structures and practices that maximise the learning opportunities

Outcome 8: Knowledge about equity in, and accessibility of, educational resources and provision

Outcome 9: Knowledge about the standard of the Bhutanese education system as compared to the standard of other countries

Outcome 10: A database for studying student achievement over time

1.4 The Bhutanese Education System

The Kingdom of Bhutan is a small country with an area of 38,394 square kilometres—about the same as Switzerland. Its population is close to half a million people. Bhutan shares its border with China in the north and India in the south. Bhutan had a Buddhist monastic education system before the arrival of modern education in the early 1960s. Since then, modern education principles and practice have rapidly emerged as underpinnings for the education system in Bhutan. By the year 2010, Bhutan had 547 schools with 170,405 students (Policy and Planning Division, 2010) in contrast to 11 schools and about 400 students in the early 1960s (Department of Education, 1999). This indicates that the schools and the students have been increasing at the rates of approximately 11 schools per year and 3,400 students per annum, since the 1960s.

Over the years, the modern education system in Bhutan has evolved into seven distinct departments: Early Childhood Care and Development Education; Primary Education; Secondary Education; Tertiary Education; Adult Literacy and Continuing Education; Technical and Vocational Education; and Special Education (Department of Education, 1999). Bhutanese children begin their schooling at the age of six years, and many of them will be 16-year-olds by the time they complete Grade 10, which is also the grade at which the basic education level is defined as having been achieved. At the end of Grade 10, students sit for the national-level examinations conducted by the Bhutan Board of Examinations to qualify for the next higher grade. Students who qualify for this next higher grade continue their education until Grade 12 over a further two years, while other students either pursue technical and vocational courses or seek a career. At the end of Grade 12, students sit for another national-level examination conducted by the Bhutan Board of Examinations. Some students qualify for tertiary education, while others enter the world of work after completing Grade 12.

This study aims to develop a model that is capable of generating indicators that may offer insights into such issues as the career prospects of students who fail to advance their schooling beyond the basic education level and the capacity of the students who leave schooling with the basic education level certificate to compete for on-the-job training or for other out-of-school job-related training.

1.5 Educational Problems and Challenges in Bhutan

The rapid progress of its modern education system, with large annual increases in schools and students, has been confronting Bhutan with numerous educational challenges. These include post-school further education enrolment pressure, misalignment between education outcomes and employment qualifications and needs, a pronounced rural-to-urban population drift, and achievement of equitable access to learning opportunities (Department of Education, 1999).

High enrolment rates and low dropout rates of students in primary and secondary schools have led to enormous pressures on teaching and learning resources across the Bhutanese education system. The resource constraints are also distinctively visible in the tertiary education sector where learning opportunities are restricted by a conventional third world “educational pyramid”. For example, owing to resource constraints, the eligibility criteria for undergraduate courses are set very high. While the eligible candidates pursue higher education, the ineligible, by virtue of lesser performance scores, which may be only marginally different, miss the opportunity and increase the likelihood that they will not reach their potential to contribute fully to the nation.

The gap between the expectations of school and university graduates and the real demand for their knowledge and skills in the world of work, that is, the number of high-status jobs generated by the economy, has led to an increasing number of unemployed youth in the country. Vocational and technical education as an alternative educational pathway has not been able to motivate school graduates, due to its perceived low status, poor working conditions, and the low salary of "blue-collar" jobs.

Rapid urbanization, particularly over the last ten years, has led to the erosion of Bhutanese traditional family ties and of the related community-based social support system, making its youth vulnerable to anti-social behaviour. Further, the rugged mountainous terrain and the widely dispersed population have created inequitable access to learning opportunities across the country, which are not easy to redress.

Clearly, the Bhutanese education system needs reform and increased relevance if it is to meet the specific human resource and development requirements of the country in a globalized economy. The international benchmarks that this study aims to generate through a national educational assessment model will provide indicators for reforms of both policy and practice.

1.6 Bhutan's Educational Problems and Challenges From a Global Perspective

The problems and challenges faced by the Bhutanese education system resemble the longstanding "ten critical future issues" identified by Coombs (1982):

- Keeping pace with rapidly *expanding learning needs* that cut across all segments of the population and society;
- Coping with a growing financial squeeze, caused by rising educational costs pressing against tighter budget ceilings;
- Rectifying serious maladjustments between education and employment;
- Overcoming unacceptable socio-economic, sex and geographic educational disparities and inequalities;
- Improving educational quality and the fitness of education to changing environmental conditions and the realistic needs and interests of learners;
- Harmonizing education and culture in each society;
- Achieving more efficient and effective use of limited educational resources;
- Developing broader, more flexible and community-based approaches to educational planning and management;
- Building a progressively broader and more diversified 'learning network'- combining formal, non-formal and informal modes of education to serve the evolving life-long learning needs of all members of the population;
- Revitalizing and reorienting international educational co-operation to meet changing needs and world conditions. (pp. 145-146)

To address these challenges, Coombs (1982) emphasized the need to develop an education system that embraces the life-long learning needs of the whole society and all modes and types of learning. Knapper and Cropley (2000) and the OECD (2004c) note that life-long education has emerged out of the need to keep abreast with unpredictable demand for, and supply of, knowledge and skills and learning opportunities because of a rapidly changing and an emerging ICT-connected world.

A vision document of the Royal Government of Bhutan, where the perspective extends to the year 2020, envisions a life-long education system in Bhutan. It states that Bhutan must have a dynamic education model that provides for “multiple entry and exit points to a variety of courses and learning opportunities that go beyond the traditional boundaries set by existing institutions” (Planning Commission, 1999, pp. 20-21). In line with the vision document, the Ministry of Education (Department of Education, 1999) has set life-long learning as one of its goals:

> The rapid globalisation of the world economy and increasing access to information has offered a new dimension for learning. One component of this new dimension is the concept of life-long learning. To that end, secondary schools must equip young people with the interest and tools needed to continue to learn and stay abreast of technological developments throughout their lives. (p. 21)

Further, a recent planning document, *Draft Tenth Five Year Plan (2008-2013) Vol. 1: Main Document* states that “the Royal Government will place a strong emphasis on promoting life-long learning through all formal and non-formal education processes as well as through effective human resource development in the public and private sectors” (GNH Commission, 2008, p. 49). All this indicates that Bhutan has been embracing a life-long education system since 1999. In addition, the planning document commits the education system to creating awareness of the benefits of a life-long education by fostering a learning culture among all sections of Bhutanese society. This is all the more relevant to Bhutan, given the 54% adult literacy rate (Royal Government of Bhutan, 2005, p. 47), 1.4% of the total labour force with a college-level education (GNH Commission, 2008, p. 48), and a 6 to 7% annual economic growth forecast (Royal Government of Bhutan, 2005, p. 23). The emphases in the policy documents indicate that Bhutan recognises the need to provide its people with learning opportunities through life-long education as a means to raising its literacy rate and the education level of its current workforce. The proposed model in this thesis will contribute to measuring the preparedness of Bhutanese students’ for life-long education.

Drawing on the works of Tuijnman, Kirsch, and Wagner (1997), Resnick (1987), and the Secretary of Labour’s Commission on Achieving Necessary Skills (1991), Knapper and Cropley (2000) identified key attributes of the concept of knowledge in the context of life-long learning. These attributes are: collecting, analysing and organizing information; communicating ideas and information; planning and organizing resources; understanding and designing systems; solving problems; using technology; using mathematical ideas and techniques; adaptation skills; analytical skills; working with others; higher-order-thinking skills, and meta-cognition. Further, Cropley (1981) and Knapper and Cropley (2000) emphasized that life-long learning also depends on non-cognitive factors such as motivation, attitudes, values, and the self-image of students because these factors define the students' readiness for life-long learning. With reference to these attributes of life-long learning, Cropley (1981) characterized a life-long learner as someone who understands the relationship between learning and real life, recognises the need for life-long learning, and possesses a positive self-concept and skills for life-long learning. Cropley (1981) also presented a set of skills necessary for life-long learning. These skills included goal setting, research skills, self-monitoring of learning, and effective use of learning devices and resources.

The model for a national educational assessment in Bhutan, as proposed in this thesis, will emphasise these aspects of knowledge and skills as broad indicators of Bhutanese students' preparedness for successful participation in the world of work after completing their school education. Therefore, as one of its goals, the model will evaluate the effectiveness of the Bhutanese education system in the context of inspiring students to life-long education through Outcomes 4 and 5.

1.7 Current Educational Assessment Programmes in Bhutan

The *Education Sector Strategy: Realising Vision 2020* (Department of Education, 1999) proposed that:

- a study shall be conducted to determine the effectiveness of pre-service programmes with a view to improve the delivery of the courses;
- an assessment of the teacher competencies in key areas of curriculum shall be undertaken in 2002;
- a baseline study on the quality of education for classes VIII and X leavers shall be conducted in 2003 involving Bhutanese educators and external consultants;
- Bhutan will also participate in an international comparison study for student achievements in English, mathematics and science. (pp. 17-24)

To implement these plans, the Ministry of Education of the Royal Government of Bhutan initiated *National Education Assessment* (NEA). NEA is "a system-wide assessment programme to investigate and monitor the health of the education system" (Bhutan Board of Examinations, 2004, p. 14). The main purposes of NEA, according to the Board (2004), were to provide:

- policy-makers with information to monitor standards over time, to monitor the impact of particular programmes, and to make decisions about resource allocation;
- schools and teachers with information about whole school, class and individual pupil performance so that they can use these to make decisions about resource allocation and to support learning in the classroom;
- the national system with information that will help to compare its performance with the international standards. (p. 14)

NEA surveyed the achievements of Bhutanese students at Grade 6 in literacy and numeracy in 2004 and at Grade 10 in English and mathematics in 2006. The NEA 2004 report provided the stakeholders in the Bhutanese education system with information about the national average scores on literacy and numeracy and some potential factors related to variations in student achievement. However, the NEA 2004 report recommended that the future NEA should increase the sample size of both teachers and students to make it more representative (Bhutan Board of Examinations, 2004). This raised a serious doubt about the validity of the NEA 2004 report, because sample size influences statistical power (Eng, 2003; Lenth, 2001). Eng (2003) and Lenth (2001) note that an inadequate sample size weakens the statistical power of studies, perhaps eliciting erroneous inferences from a sample-based research study. In addition, because NEA 2004 was carried out exclusively in Bhutanese schools without any attempt to link to any of the international assessment programmes, it could not compare results with international benchmarks.

Based on the experience from the NEA 2004, the NEA 2006 adopted a census approach. However, due to resource constraints related to full-time overseas consultancy services, the utility of the NEA 2006 was not fully explored. Further, the NEA 2006 could not provide international benchmarks for the reasons similar to the NEA 2004 (Bhutan Board of Examinations, 2008). Therefore, it is clear that Bhutan has to have a valid and reliable model to guide the national educational assessment programme, both to evaluate the current quality of its school education system and provide an opportunity for international benchmarking. This thesis aims to address these needs.

1.8 Researcher's Background Relevant to the Study

I worked for the Ministry of Education of the Royal Government of Bhutan as a teacher of mathematics and science in a number of schools in Bhutan for five years. I also worked as a Principal in one of those schools in the last two years of my brief teaching career. It was during my days in those schools that I developed a keen interest in educational assessment and evaluation.

As a teacher, I practised different forms of educational assessment and evaluation (e.g., diagnostic, formative, summative). As a principal, I recall using national mean scores of Bhutanese students on the national examinations published by the Bhutan Board of Examinations. However, the types of assessments I practised in the classroom and the national examination score sheets published by the Bhutan Board of Examinations could not generate sufficient information about the factors related to student achievement that I required as a teacher and principal. In addition, the score sheets could not be used to process information about other schools to compare and exchange experiences and ideas likely to have caused the difference in student achievements among schools.

After five years of teaching in schools, I was transferred to the Bhutan Board of Examinations where I worked for four years as a Subject Specialist in mathematics and Coordinator for the National Education Assessment programme. As a Subject Specialist, my responsibility was to design and develop national examination papers for mathematics and computer science in collaboration with senior teachers. As the National Education Assessment Coordinator, my responsibility was to coordinate and conduct the national education assessment programme in Bhutan.

In this capacity as the national coordinator, I had various opportunities to work with expert consultants from the Australian Council for Educational Research in Australia and from the Dutch National Institute of Measurement in the Netherlands. I also participated in a series of capacity building workshops that were organized by the World Bank Institute in various cities, where I met the National Assessment Coordinators from other countries. It was in those workshops that I realised the potential of, and the challenges in, national educational assessment. A valid and reliable national educational assessment has great potential to monitor and guide the progress of a nation's education system. However, the psychometric tools used for processing information from the national educational assessment data are not only formidably intricate, but are also a set of procedures best explored during advanced academic studies. I was fortunate enough to be granted a NUFFIC Fellowship to pursue a M.Sc in educational evaluation and assessment at the University of Twente at Enschede in the Netherlands. My master's thesis, titled *IRT in Item Banking, Study of DIF Items and Test Construction* (Tshering, 2006), was on the use of item response theory in developing an item bank, detecting differentially functioning items and constructing tests. The research for the master's thesis was conducted at the Dutch National Institute of Educational Measurement at Arnhem (alias CITO) in the Netherlands. I was fascinated by the outcomes of the master's thesis, especially when tests were developed for Dutch students based on its findings. Further, my interactions with psychometricians at the Department of Psychometrics at CITO instilled in me a strong desire to learn more about psychometrics.

The prospect of monitoring and guiding the Bhutanese education system by using advanced psychometrics seemed promising. In recent years, the Bhutan Government has suffered frequent criticisms from the public on the quality of school education (Ministry of Education, 2006). Valid and reliable large-scale assessment programmes like *Programme for International Student Assessment* (OECD, 1999b, 2000, 2004b, 2006, 2010), *Trends in International Mathematics and Science Study* (IEA, 1998; Mullis, Kennedy, Martin, & Sainbury, 2006; Mullis, Martin, Ruddock, Sullivan, & Preuschoff, 2009; Mullis et al., 2003) and *National Assessment of Educational Progress* (NAGB, 2001, 2004, 2007a) are known for their capacity to generate evidence about various factors affecting student outcomes. Such information can also give insight into the issues related to the quality of education and facilitate informed policy-decisions related to school improvement programmes. Based on these diverse experiences, I chose to pursue a doctoral study in educational evaluation and assessment, with a special focus on large-scale educational assessments.

A valid and reliable educational assessment and evaluation system is inextricably related to the overall effectiveness of an education system. In other words, a truly effective education system will have a valid and reliable educational assessment and evaluation system embedded in it. This inextricable link also implies that the latter has to be as responsive as the former to the needs of the society on whose goals, values, and culture it is founded. My brief five-year school experience and four-year administrative experience at the Bhutan Board of Examinations of the Ministry of Education of the Royal Government of Bhutan confirm in me the notion of educational effectiveness as being a complex construct that requires an eclectic approach (integrating economics, sociology, and psychology) to fully comprehend its multiple dimensions.

I also expect that the complex but learnable psychometrics and educational effectiveness research, which underpin this doctoral thesis, will assist me in being resourceful in instituting a strong culture of research in the field of educational assessment and evaluation in the Bhutanese education system. In addition, I hope to participate in, and contribute to, research efforts in educational communities across the globe in the field of educational assessment and evaluation. The research aim outlined for this thesis set the research primarily in the Bhutanese context, but the methods and procedures used in it make the thesis relevant to assessment procedures of other similarly placed nations.

1.9 Introduction to the Research Methods

A concurrent mixed methods research approach will be used in this thesis (Creswell, 2003; Greene, Caracelli, & Graham, 1989; R. B. Johnson & Onwuegbuzie, 2004; R. B. Johnson, Onwuegbuzie, & Turner, 2007; Tashakkori & Teddlie, 1998), with a cross-sectional survey (Babbie, 1990; Creswell, 2003; Fowler, 2009; Rea & Parker, 2005) and a focus group interview (Krueger & Casey, 2009;

Morgan, 1988; Morgan & Scannell, 1998) as two data collection tools. As described in the later chapters, these methods are chosen based on the nature of the aim and outcomes of this thesis.

1.10 Significance of the Study

A national educational assessment model such as this study seeks to develop should have significant value to diverse stakeholders in the Bhutanese education system and international researchers in educational assessment and evaluation. Information processed from the national educational assessment data will be of practical use to Bhutanese stakeholders, such as students, teachers, parents, taxpayers, and policy-makers. Data-driven educational debates and policy-decisions should have considerable potential to help the stakeholders in addressing the problems of the nation's education system and in developing efficient educational interventions. In addition, a national assessment model and the methods used in developing it should be of interest to the communities of researchers in educational assessment and evaluation. A further value of the study to overseas researchers is the database that it aims to develop.

Considering the potential wide-ranging benefits and readership of the study, its significance is revisited in Chapter 11.

1.11 Limitations of the Study

Like any research study, this study has a number of limitations. First, the study is cross-sectional: that is, it is conducted as a one-off PhD study, and therefore the study cannot measure causality between variables. Second, the questionnaires used in the study are self-administered by the participants. As with self-report data, survey questionnaire results are subjective and they differ from objectively measured data. Third, the lack of sufficient funding for field research also limited the number of participants, and the implication is especially prominent in the qualitative part of the study. These limitations are revisited in Chapter 11 where their implications to the validity of the study are examined critically.

1.12 Chapter Summary

This chapter has presented a context for the study by examining the need for a national educational assessment model in Bhutan. Bhutan's initiative to make schools more autonomous, set international benchmarks, follow standards-based school curricula, use evidence-based research, and develop an assessment database provide a suitable context for the use of a national educational assessment model. Drawing on this contextual analysis, ten outcomes have been presented for the model [see Section 1.3], that also reflect success indicators for this study.

A brief review of the history of the Bhutanese education system indicates that Bhutan's school education system has been expanding rapidly, presenting it with numerous challenges such as enrolment pressure on higher education institutes, maladjustment between school curricula and employment, and rural-urban drift. The analyses of Bhutan's policy documents have shown that Bhutan recognises life-long education as a measure to counter the challenges related to its education system. It has been proposed that the system-wide educational assessment model that this study seeks to develop will have the capacity to measure the preparedness of Bhutanese students for life-long learning.

It has been indicated that the study will use concurrent mixed methods as the research method, with cross-sectional survey and focus group interview as the data collection instruments. In addition, some possible strengths and limitations of this study have also been indicated.

1.13 Chapter Conclusion

Critical analyses of historical developments, challenges, and policy directives of the Bhutanese education system pointed to life-long education as a prominent education model that Bhutan has been pursuing over time. Further, an analysis of the Bhutanese education system showed that Bhutan requires a national educational assessment model capable of providing it with comprehensive information about the health of its education system. Therefore, to develop a national educational assessment model for Bhutan has been defined as the aim of this study. In addition, it has been indicated that the model should have the capacity to supply Bhutan with information underpinning the 10 outcomes that have been formulated for this research [see Section 1.3].

1.14 Thesis Overview

Chapter 2 reviews some globally reputed large-scale assessment frameworks, with emphasis on their principles, research questions, design methodologies, contents, grades assessed, validity issues, and educational impacts. In addition, the chapter identifies a set of broad principles, based on their compatibility with the Bhutanese education system, for use as guidelines in developing a national educational assessment model for the Kingdom of Bhutan. These principles are used to devise a large-scale assessment framework for the Bhutanese education system in the later chapters.

Chapter 3 reviews different traditions of educational effectiveness research, with a view to developing a conceptual national educational assessment model for use in this thesis. Drawing on NAEP, TIMSS, and PISA, the chapter examines critically the various contextual variables identified in the literature on educational effectiveness. Finally, the chapter concludes by presenting a national educational assessment model that takes account of educational effectiveness factors relevant to Bhutan and other education systems.

Chapter 4 reviews a range of research designs with potential for use in the proposed national educational assessment model, presents the most appropriate methods for the model, and shows how research instruments are developed for use in this thesis. Also discussed in the chapter are the field administrations of the research instruments, including validity issues, and the preparation of data for analyses in the later chapters.

Chapter 5 presents the results from the analyses of data from the Mathematics test based on the following analytical framework: (a) describing data analysis procedures, (b) presenting statistical indices, (c) relating research objectives to statistical indices, and (d) informing policy perspectives. Guided by this analytical framework, the chapter links the Mathematics test to PISA's Mathematical Literacy test, profiles students' mathematical knowledge and skills, compares schools, sets international benchmarks, and discusses policy implications.

Chapter 6 reports on the results from the analyses of data from the student questionnaire, and identifies key non-cognitive variables that relate to students' performance on the Mathematics test. The analyses are done in line with the analytical framework followed in Chapter 5. The contextual and non-cognitive variables are analysed in the order in which they were discussed in Chapter 3.

Chapter 7 presents the results from the analyses of data from the teacher questionnaire and identifies the key teacher-related variables that relate to student performance on the Mathematics test. The analyses are done in line with the analytical framework presented in Chapter 5.

Chapter 8 reports on the results from the analyses of data from the school questionnaire and identifies the key school characteristics that relate to students' performance on the Mathematics test. The analyses were done in line with the analytical framework presented in Chapter 5.

Chapter 9 presents the results from the analyses of data from the focus group interview. The chapter begins by reporting the situation in which the focus group interview is conducted. The chapter then presents critical analyses of the focus group interview questions. The chapter also presents policy implications from the findings, followed by some pertinent conclusions.

Chapter 10 consolidates all the sectional discussions presented in Chapters 5 through 9, drawing together the emerging themes, and discussing strategies for disseminating the findings from the study to stakeholders in the Bhutanese education system and to the communities of researchers in educational assessment and evaluation.

Chapter 11 presents conclusions to the study, revisits the significance and the limitations of the study, makes some recommendations from the study and suggests potential areas for further research.

Chapter 2

LARGE-SCALE ASSESSMENTS AND THEIR FRAMEWORKS

2.1 Introduction

This chapter aims to examine critically three internationally reputed large-scale assessment frameworks: NAEP, TIMSS, and PISA. In addition, the chapter identifies a set of broad principles, based on their compatibility with the Bhutanese education system, for use as guidelines in developing a national educational assessment model for the Kingdom of Bhutan.

The chapter compares the three assessment programmes, presents current views on large-scale assessments, and compares large-scale assessments and high-stakes examinations. The chapter also consolidates certain design features of the three assessment programmes for adaptation to a national educational assessment model for Bhutan.

2.2 Large-Scale Assessments

This section reviews three well-established large-scale assessment frameworks in terms of their principles, research questions, design methodologies, contents, grades assessed, validity issues, and educational impacts. The critical review is expected to guide the development of a national educational assessment model for Bhutan.

2.2.1 National Assessment of Educational Progress (NAEP)

As asserted by the National Assessment Governing Board (NAGB), NAEP, introduced in 1969, is a Congressionally mandated project of the United States's Department of Education's National Centre for Education Statistics (NAGB, 2003). Since then it has been assessing the students of the United States in geography, reading, writing, mathematics, science, the history of the United States, the arts, civics, and other academic subjects. NAEP also collects information about the characteristics of students, teachers, and schools that are known to influence student achievement (NAGB, 2003). Unlike early NAEP stages, where students were sampled by their age, the present NAEP samples students by their grades, with the grades being 4, 8, and 12 (NAGB, 2005, 2006a). NAEP is administered every two years to a nationally representative sample of students from public and private schools across the United States.

The features of NAEP such as the framework definition, the research objectives, the design elements, and the mathematics frameworks are examined critically in the next section. In addition, some of the influences of NAEP on the American school education system are also discussed.

2.2.1.1 Definition of the NAEP Frameworks

The NAGB (2005, p. 2) notes that the NAEP frameworks are designed to generate outcomes based on what is deemed as "essential learning" in specific subjects to set the stage for assessments. Drawing on this general purpose of the NAEP frameworks, the NAGB (2007a), defines an assessment framework as follows:

> An assessment framework is like a blueprint. It lays out the basic design of the assessment by describing the mathematics [and science] content that should be tested and the types of assessment questions that should be included. It also describes how the various design factors should be balanced across the assessment. (p. 2)

This definition makes the NAEP frameworks somewhat comparable to a conventional test blueprint, though more complex than many such examples. For instance, each NAEP framework outlines the NAEP research objectives, context for planning assessments, design principles, achievement levels, subject content areas, complexity of items, distribution of items, item formats, item scoring guides, and resources required for assessments.

2.2.1.2 The NAEP Objectives

NAEP has three broad objectives (NAGB, 2005, 2006a). First, NAEP seeks to measure national and state educational progress toward the third National Education Goal ("Archived: GOALS 2000: Educate America Act,") and provide timely, fair, and accurate data about student achievement at the national level, across the states, and in comparison with other nations. Second, NAEP seeks to develop, through a national consensus, sound assessments to measure what students know and can do as well as what students should know and be able to do. Third, NAEP seeks to help states and others to link their assessments with it and to use its data to improve educational performance.

Guided by the above objectives, NAEP reaches out to its stakeholders through a range of purposes as emphasized in the NAGB (2006a). NAEP informs the United States citizens, the curriculum specialists, and the policy-makers about the level and nature of students' comprehension of school subjects and the factors related to schooling and their association with student performance. NAEP is also lauded for providing its stakeholders with comparative student data according to race, ethnicity, type of community, and geographic region. In addition, NAEP provides its stakeholders with trends in student performance, reports on relationships between student achievement and certain background variables, and presents information on strengths and weaknesses in students' understanding and their ability to apply that understanding to problem-solving situations. These features of NAEP are consistently emphasized in a range of reports (Lindquist, 1989; NAGB, 1996, 2001, 2006b; National Center for Education Statistics, 2011), indicating their significance to the American school education system.

2.2.1.3 The NAEP Framework and Its Design Elements

Three design elements such as the national consensual approach, accommodation of views of diverse stakeholders, and assessment of student, teacher, and school pervade the NAEP framework design (NAGB, 1996, 2005, 2006a, 2007a, 2007b).

NAEP is governed by the NAGB that formulates policies, oversees the designs of assessment objectives and test specifications, researches assessment methodologies, identifies appropriate achievement goals for each age and grade, and carries out other NAEP policy responsibilities (NAGB, 2005, 2006a). The NAGB solicits advice from its committees such as the Framework Steering Committee (FSC) and the Framework Planning Committee (FPC) on its responsibilities. The FSC and the FPC collaborate with each other, with the former developing principles that underpin the NAEP frameworks and providing policy guidelines to the latter. The FPC, which is responsible for identifying goals and objectives and formulating assessment frameworks for NAEP, consists of scientists, practitioners, and recognized experts in the subjects assessed by NAEP.

NAEP is designed consensually with active participation of curriculum specialists, subject teachers, local subject supervisors, state supervisors, administrators, parents, scientific associations, business and industry, government, unions, and psychologists (NAGB, 2005, 2006a). Through this consensual approach, NAEP has been able to embody a range of views, experiences, and expertise, resulting in greater ownership among its stakeholders. The NAGB (2005, 2006a) states conclusively that it uses well-formulated communication strategies, well-defined subject scope, and research-based claims and experiences to accommodate the views of its diverse stakeholders in the NAEP framework.

NAEP assesses a nationally representative sample of students from public and private sector schools in Grades 4, 8, and 12 every two years. For example, in the 2005 NAEP science assessment, 8,500 schools and 147,700 students participated in Grade 4 NAEP science assessment and 6,400 schools and 143,400 schools participated in Grade 8 NAEP science assessment (NAGB, 2006b). Therefore, NAEP is a large-scale, cross-sectional survey that is timed, standardized, and low-stakes for students, teachers, and administrators.

NAEP uses Balanced Incomplete Block (BIB) design, a variant of matrix sampling, as the item sampling technique (E. G. Johnson, 1992). BIB minimizes student burden and maximizes content coverage by increasing the number of test items. For sampling students, NAEP uses a multi-stage probability sampling technique.

2.2.1.4 The NAEP Frameworks

NAEP has different assessment frameworks for different subjects. However, they have generic elements across school subjects. The elements characteristic of a NAEP framework are content areas, complexity of items, item formats, item distribution, item scoring guides, achievement levels, assessment duration, accommodation for special needs students, and adaptation for limited English speakers (NAGB, 2005, 2006a).

The salient features of the NAEP frameworks for mathematics for Grades 4 and 8 are described briefly in the next section.

2.2.1.5 The NAEP Mathematics Framework for Grades 4 and 8

The NAEP mathematics framework has two dimensions: content areas and mathematical complexity of items. In addition, the framework has design elements like item distribution, item formats, manipulative, calculators, achievement levels, and accessibility options (NAGB, 2006a). These features are summarized in Table 2.1.

Table 2. 1. Structural Features of the NAEP Mathematics Framework

Dimensions	Attributes	Objectives
Content	• Number Properties and Operations • Measurement • Geometry • Data Analysis and Probability • Algebra	To find out what students know and can do in mathematics
Mathematical Complexity of items	• Low complexity • Moderate complexity • High complexity	To identify the complexity of demands an item makes on a student’s thinking
Additional Design Elements	Attributes	Objectives
Item Distribution	• Items are distributed across content areas in percentages	To reflect the importance and value given to each of the curricular content areas within mathematics
Item Formats	• Multiple choice • Short constructed response • Extended constructed response	To accommodate different content areas and skills with appropriate items
Manipulative	• Specific manipulative not available	To facilitate authentic assessment

Calculators	• Four function calculator • Scientific calculator	To identify items for use of calculators and provide Grade specific calculators
Achievement Levels	• Basic • Proficient • Advanced	To provide descriptions of what students know and can do in mathematics
Accessibility	• One-on-one testing • Small group testing • Extended time • Oral reading of directions • Large-print booklets • Bilingual booklets • Use of an aide to transcribe responses	To maintain equity and universal accessibility amidst students of varying backgrounds and with special needs

The elements like manipulative operations and accessibility are not prominent in other large-scale assessments. However, the NAEP framework clearly cautions about the use of manipulative materials because of their implications on cost, time, and invigilation, and suggests that manipulative materials should be reasonable, should be used where possible, and should not disrupt the test administration process (NAGB, 2007a). The accessibility element of the NAEP framework offers an unbiased opportunity for the students with special needs to demonstrate their knowledge and skills.

2.2.1.6 NAEP and Its Impacts and Critical Responses

This section discusses NAEP's influence on the American school education system and examines its limitations.

NAEP has been well-received by its stakeholders in America. Daro, Stancavage, Ortega, DesStefano, and Linn (2007, p. i) state that NAEP "is the only source of information on the educational attainment of all United States's students, and it is the only vehicle by which states can compare the progress of their students against a common standard". A multitude of secondary research articles based on NAEP data also demonstrate the penetration by NAEP into research communities (Abedi, Lord, & Hofstetter, 1998; Beaton & Gonzalez, 1995; Daro, et al., 2007; E. G. Johnson & Owen, 1998; Lindquist, 2001; G. W. Phillips, 2007; Reckase, 2002). Bourque (1999) notes that NAEP data have been informing many national policy debates, and that these data have been an integral part of the American national evaluation of educational progress at national, state, and local levels since 1969. Fuller, Gesicki, Kang, and Wright (2006) trace the use of evidence from NAEP data by George W. Bush, the then President of the United States, in his presidential campaign to support his claim of the progress made by America's children under his presidency, indicating the vast significance of NAEP to the American school education system. Bourque (1999) sums up the significance of NAEP to the United States school education system by noting it as America's national barometer of educational achievement.

NAEP, like any other study, has its limitations. The performance standards of earlier NAEP had been the most critiqued standard-setting effort since its conception [(Bourque, 1999), cf. ACT, 1993; NAE,1993a, 1993b,1994; United States GAO, 1993; NAGB & NCES, 1995]. The contradiction between defining anchor points and descriptive achievement levels was regarded as highly controversial [(Bourque, 1999), cf. ACT, 1993; NAE,1993a, 1993b,1994; United States GAO, 1993; NAGB & NCES, 1995]. It had been argued that anchor points were set quantitatively, while descriptive achievement levels were set qualitatively. NAEP reconstituted the committees to set reliable and valid achievement levels in response to the critiques. Consequently, the current NAEP achievement levels are reported to have passed a legislated national consensus process as measures for their reliability and validity (NAGB, 2000). The revised NAEP achievement levels include all ranges of the performance distribution.

The information processed from the earlier version of NAEP that used an age-based sample of students was criticised for being vague and confusing. Because schools are organized around grade structure, not age, the information based on an age-based sample of students was fraught with confusion when it was used in guiding and informing programme decisions at the appropriate instructional levels (Bourque, 1999). As a result, the current version of NAEP applies grade-based samples of students.

Fuller, Gesicki, Kang, and Wright (2006) noted that NAEP tends to influence the curriculum standards of the states when it is associated with accountability. This risks narrowing the curriculum by the states to match what is assessed by NAEP, resulting in less teaching and learning in schools despite high scores on NAEP. Further, Fuller et al. (2006) note that the scores from state assessments, when compared with the NAEP scores, were generally inflated over time and that their cut-off points were set lower than that of NAEP. This underscores the tendency of the states to inflate scores or water down the curriculum standards by setting low cut-off points. To avoid such inclinations and actions, NAEP is profiled as low-stakes assessment for students, teachers, and schools.

Fuller et al. (2006) expressed reservation about the use of NAEP scores to infer discrete effects of federal 'No Child Left Behind Policy' reforms, because NAEP is not aligned to any particular state's curricular standards. This implies that NAEP does not always test what is actually taught in schools. However, a validity study of NAEP concluded that "the NAEP Mathematics Framework and accompanying specifications provide a reasonable representation of the domain of fourth- and eighth-grade mathematics when compared to states and other nations" (Daro, et al., 2007, p. 124). These debates demonstrate the need to study the potential gap between school curricula and large-scale assessment frameworks to make the findings from the latter relevant to the former, and this will be considered in the design of the proposed model for Bhutan.

The information processed from NAEP is disaggregated at student, state, and national levels (Kafer, 2001). However, as NAEP uses random samples of students from participating states, its information cannot show whether a particular student is reading proficiently or how his or her school compares to other schools (Fuller, et al., 2006). In addition, it has been noted that NAEP's lack of individual achievement measures from testing on more than one occasion is its biggest limitation for drawing causal inferences, particularly about effects on individual students (Walberg, 2002).

In their study of the validity of NAEP mathematics assessments for Grades 4 and 8, Daro et al. (2007) observed that the NAEP mathematics assessment is situated "behind" the framework but "ahead" of the population, implying that it has failed to capture the challenging contents of its framework. To address this shortfall, "NAEP has been recommended to encompass the achievement of the full population—from the lowest to the highest—and reach from the least to the most advanced content of the frameworks" (Daro, et al., 2007, p. 119). Further, Daro and associates suggest the inclusion of more easy items, more high-complexity items, and more items that reach forward in the curriculum. These observations highlight the challenges involved in designing a fair assessment instrument with competing priorities such as the content coverage, the item complexity, and the candidates' capabilities. Notwithstanding the criticisms, Daro et al. (2007) assert that NAEP is sufficiently robust to support the main conclusions that have been drawn about the United States and its states' progress in mathematics since 1990.

In summary, a promising inference from the review of NAEP is that it has been instrumental in driving educational debates and interventions in the United States, which is also the goal of a system-wide educational assessment model that this study seeks to develop for Bhutan. This makes NAEP an ideal point of departure for developing a system-wide educational assessment model.

2.2.2 Trends in International Mathematics and Science Study (TIMSS)

TIMSS is a project of the International Association for the Evaluation of Educational Achievement (IEA). Since its creation in 1959, TIMSS has grown in size to have 59 countries with approximately 425,000 student participants in 2007 (Husen & Postlethwaite, 1996; Mullis et al., 2008; Mullis et al., 2005). Currently, IEA has a secretariat at Amsterdam in the Netherlands. TIMSS assesses students in science and mathematics at Grades 4, 8, and 12 and links student achievement to student, teacher, and school characteristics that are known to relate to student achievement.

2.2.2.1 Definition of the TIMSS Framework

TIMSS depicts implicitly an assessment framework as a document that specifies assessment objectives, curriculum content, test formats, and the overall purposes of assessment (Mullis, et al., 2005).

The TIMSS framework describes the TIMSS curriculum model, documents development procedures for the framework, defines research questions, and specifies subject content and cognitive domains (Mullis, et al., 2006; Mullis, et al., 2005; Mullis, et al., 2009; Mullis, et al., 2003). Additionally, the TIMSS framework describes assessment design elements like item formats, achievement levels, scoring guides, and assessment resources. These features of the TIMSS framework are similar to the features of the NAEP frameworks, indicating a general consistency in their assessment framework designs.

2.2.2.2 The TIMSS Research Objectives

The TIMSS's research objectives are implicit in its purposes. According to Mullis et al. (2005), the countries that participate in TIMSS can:

- have comprehensive and internationally comparable data about what mathematics and science concepts, processes, and attitudes students have learned by the fourth and eighth Grades;
- assess progress internationally in mathematics and science learning across time for students in Grades 4 and 8;
- identify aspects of growth in mathematical and scientific knowledge and skills from fourth to eighth Grades;
- monitor the relative effectiveness of teaching and learning at the fourth Grade as compared to the eighth Grade, since the cohort of fourth-Grade students is assessed again as eighth-Graders;
- understand the contexts in which students learn best. TIMSS enables international comparisons among the key policy variables in curriculum, instruction, and resources that result in higher levels of student achievement; and
- use TIMSS to address internal policy issues. Within countries, for example, TIMSS provides an opportunity to examine the performance of population subgroups and address equity concerns. It is efficient for countries to add questions of national importance (national options) as part of their data collection effort. (p. 10)

These purposes feature in a series of TIMSS reports (Mullis, Kennedy, Martin, & Foy, 2007; Mullis, et al., 2008; Mullis et al., 2000a; Mullis et al., 2000b; Mullis, Matin, Gonzalez, & Chrostowski, 2004), together with complete technical know-how of processing the TIMSS data in technical reports (IEA, 2000, 2004, 2008), providing a rich knowledge base for developing new large-scale assessments such as a system-wide educational assessment model that this study aims to develop.

2.2.2.3 The TIMSS Framework and Its Design Elements

As with NAEP, TIMSS is developed by committees and organizations by using consensual approaches and diverse expertise. TIMSS is managed by TIMSS and Progress in International Reading Literacy Study (PIRLS) Study Center at Boston College, United States. The Centre operates through various groups like the Science and Mathematics Item Review Committees, the International Expert Panel, the National Research Coordinators, and the National Committees. These groups assist the TIMSS and PIRLS Study Centre in designing assessment frameworks, writing assessment items, and matching national curricular priorities of the participating countries using TIMSS. The TIMSS and PIRLS Study Center also works closely with other organizations like the IEA Data Processing Center in Hamburg on database creation and documentation, the Statistics Canada in Ottawa on sampling, and the Educational Testing Service in Princeton, New Jersey on the psychometric scaling of data.

The TIMSS framework is based on its curriculum model that categorizes a curriculum into three broad categories: the intended curriculum, the implemented curriculum, and the achieved curriculum (Mullis, et al., 2005; Mullis, et al., 2003). The intended curriculum represents mathematics and science that society intends its students should learn, and how the education system should facilitate this learning. The implemented curriculum represents what mathematics and science is actually taught in classroom, who teaches this content, and how it is taught. The achieved curriculum represents the mathematics and science that students have actually learnt, and how they think about these subjects.

TIMSS is a cross-sectional survey with student, teacher, and school questionnaires as the survey instruments (Mullis, et al., 2005). TIMSS uses matrix sampling and stratified random sampling to sample test items and students respectively.

2.2.2.4 The TIMSS Mathematics Framework

The TIMSS mathematics framework is organized around two dimensions: the content and the cognitive dimensions. The content dimension details the subject matter to be assessed within mathematics. The cognitive dimension specifies the sets of behaviours expected of the students as they learn the mathematics content. The dimensions are defined by a number of domains that are again designed to meet several objectives (Mullis, et al., 2005). Table 2.2 (Mullis, et al., 2005) shows the two dimensions of the TIMSS 2007 Mathematics Framework for Grades 4 and 8.

Table 2. 2. Structural Features of the TIMSS 2007 Mathematics Framework

Grade	Dimension	Domains
4	Content	• Number • Geometric Shapes and Measures • Data Display
	Cognitive	• Knowing • Applying • Reasoning
8	Content	• Number • Algebra • Geometry • Data and Chance
	Cognitive	• Knowing • Applying • Reasoning

The categories in the cognitive domain of mathematics are characterized by behaviours such as recall, recognize, compute, retrieve, measure, and classify to demonstrate "Knowing"; select, represent, generalize, synthesize or integrate, justify, and solve routine problems to demonstrate "Applying"; and analyse, generalize, synthesize/integrate, justify, and solve non-routine problems to demonstrate "Reasoning" (Mullis, et al., 2005). These skills form a continuum with recall and evaluation at its two ends, ensuring that TIMSS covers the appropriate range of cognitive skills across its content domains.

2.2.2.5 TIMSS and Its Impacts and Critical Responses

TIMSS has been instrumental in fostering national and cross-national educational debates, driving educational policy decisions, and guiding educational reforms. This section presents an overview of TIMSS's impacts on the school education systems of the seven of its participant countries.

In Hungary, consistently poor achievement results in reading comprehension from TIMSS 1979, 1980, and 1986 shocked the Hungarian Ministry of Education with sharp reactions from both education professionals and the public (Bathory, 1992). The experience compelled Hungary to institute various interventions in teaching methods.

In Japan, results from TIMSS led to criticisms of the Japanese education system that provided students with a narrow teacher-centred, recall-based style of instruction that produced students with poor skills and weak attitudes for inquiry learning. (Watanabe, 1992). These findings prompted Japan to various reforms emphasizing observation, experiment, and inquiry-based approaches to teaching and learning to foster scientific thinking skills in students. In addition, Japanese school teachers are reported to be using information from TIMSS's national report to diagnose problems in teaching and learning in order to improve instruction (Watanabe, 1992).

In Kuwait, TIMSS helped to promote assessment research through new survey technologies and methodologies and to monitor student achievement at primary, intermediate, and secondary levels of schooling (Hussein, 1992). TIMSS also helped Kuwait in designing frameworks and procedures for writing test items and interpreting and analysing the results and findings of similar studies in the future.

In Finland, the information provided by TIMSS was used to improve its education system at different organizational levels (Leimu, 1992). The Finnish government used the information from TIMSS to initiate and develop national-level policy discussion about the state of its education system. Finnish teachers used the information from TIMSS to comprehend the complex link between non-cognitive factors and student achievement, and Finnish researchers used the TIMSS's data and models for secondary research in teaching and learning.

In the U.K. results from TIMSS 1996 compelled the nation to look to the East for directions to raise its educational performance (Elliott, Hufton, & Illushin, 2000). Classroom practices in Eastern Europe and Taiwan were found to be a key to maintaining high standards of achievement, and this forced the U.K. to reconsider the classroom organization and pedagogical approaches practised in its schools (Elliott, et al., 2000).

In South Africa, TIMSS 1996 sparked a parliamentary debate that compelled its Minister of Education to outline certain priority areas in education. The debate resulted in highlighting some possible areas of interventions such as matching South Africa's school education curriculum with international curricula, improving the poor quality of South Africa's school text-books, and encouraging innovative methods and classroom approaches in South African schools [(Howie, 2000), cf. National Assembly, 1997, pp.111-114]. Further, the debate prompted a report to the Parliament on the need to acquire information on the syllabi and methodologies of at least seven high-performing countries for comparison, with a view to developing a national mathematics, science, and technology curricula of higher quality for South African schools [(Howie, 2000), cf. National Assembly, 1997, pp.111-114].

In Australia, TIMSS led to a call from senior university mathematicians for action to be taken to improve the mathematical skills of Australian students at primary and secondary levels (Lokan, 2000).

In summary, the various reactions to the TIMSS findings suggest that TIMSS has been instrumental in fostering educational debates and reforms in its participant countries, which is also one of the aims of a national educational assessment model that this study seeks to develop for Bhutan.

2.2.3 Programme for International Student Assessment (PISA)

PISA is an assessment programme developed by the Organization for Economic Co-operation and Development (OECD) based in France. Since its creation in 1997, the number of participating countries has been on the rise. Sixty-five countries with approximately 470, 000 students participated in PISA 2009 (OECD, 2010). PISA assesses students' reading, mathematical, and scientific literacy at age 15. Additionally, it assesses student and school characteristics that are known to be associated with student learning.

2.2.3.1 Definition of the PISA Framework

As stated in the OECD (1999b), the PISA framework is underpinned by the following design principles:

- a framework should begin with a general definition or statement of purpose–one that guides the rationale for the survey and for what should be measured;
- a framework should identify various task characteristics and indicate how these characteristics will be used in constructing the tasks; and
- variables associated with each task characteristic should be specified, and those that appear to have the largest impact on the variance in task difficulties should be used to create an interpretative scheme for the scale. (p.18)

Underpinned by these guiding principles, the PISA framework has subject domain definitions, theoretical bases for the framework, descriptions about organization of subject domains, descriptions of content domains and competences involved, situations and contents for every test item, details about how items are developed and scored, and descriptions of various reporting scales (OECD, 1999b, 2000, 2004b, 2006). These features are similar to the features noted for NAEP and TIMSS frameworks above, indicating a considerable degree of consistency between the frameworks.

2.2.3.2 PISA Research Questions

PISA is firmly rooted in the belief that parents, students, and other stakeholders should know the answers to the following broad questions about the quality of their education system (OECD, 1999b, p. 7; 2004d, p. 3; 2007a, p. 16): "Are students well prepared to meet the challenges of the future? Are they able to analyse, reason, and communicate their ideas effectively? Do they have the capacity to continue learning throughout life?" While these questions are focussed on students, PISA also collects information about schools' organizational structures and resources (OECD, 2006). Drawing on these questions and other PISA-related publications, Thomson (2008) convincingly summarises PISA's purposes in the following questions:

- How well are young adults prepared to meet the challenges of the future?
- What skills do the young adults possess that will facilitate their capacity to adapt to rapid societal change?
- Are some ways of organizing schools and school learning more effective than others?
- What influence does the quality of resources have on student outcomes?
- What educational structures and practices maximise the opportunities of students from disadvantaged backgrounds? and
- How equitable is education resource provision for students from all backgrounds?

(p.2)

These research questions show that PISA is designed to assess students not only in what they are supposed to learn in school, but also their ability to apply the knowledge and skills, acquired through schooling, to unfamiliar situations in real life (OECD, 2004b, 2006; Schleicher, 2000). In addition, Harlen (2001) and Orpwood (2000) state authoritatively that PISA contributes to a common view amongst the participating countries about what their education systems should provide to prepare their future citizens for adult life and for life-long learning, and that the countries rely on PISA's capability to assess education systems in line with the tenets of life-long education.

The research questions also encompass assumptions about transfer of learning, social justice, relevance of knowledge and skills, effective pedagogy, and future-oriented education. Implicit in the research questions are also other procedures such as international benchmarking, study of non-cognitive variables to account for differences in student achievement, and analyses of policy decisions and their impacts on education at a systemic level.

2.2.3.3 The PISA Framework and Its Design Elements

Similar to NAEP and TIMSS, PISA demonstrates a consensual approach, diverse expertise, matrix-item sampling, and multi-stage probability sampling as some of its framework design elements.

PISA is steered by a Board of Participating Countries (BPC) whose responsibilities are to determine the policy priorities for PISA and to ensure the compliance of PISA procedures with the policy priorities. PISA has a Technical Advisory Group that is responsible for all matters relating to data collection and analyses. PISA also has other groups such as a Subject Expert Group and a Cultural Review Panel with responsibilities to look after its domain specific and contextual issues. The design and implementation of PISA, approved by the BPC, is overseen by an international consortium led by the Australian Council for Educational Research (ACER) based in Melbourne, Australia. Other partners in the consortium are the National Institute for Educational Measurement (CITO) in the Netherlands, WESTAT and the Educational Testing Service (ETS) in the United States, and the National Institute for Educational Policy Research (NIER) in Japan. The Board is eventually responsible to the OECD Secretariat. PISA is implemented in the participating countries through their National Project Managers in accordance with the administration procedures formulated by the BPC. With such organizational structures, PISA is able to claim test construct validity, accommodate cultural and curricular contexts of participating countries, ensure test items of strong measurement properties, and make persuasive claims about its authenticity and educational validity of its procedures and actions (OECD, 1999b, 2004b, 2006).

PISA has three domains of assessment: reading literacy, mathematical literacy, and scientific literacy (OECD, 1999b, 2004b, 2006). PISA is conducted in a three-year cycle with one of the domains identified as the major domain and the other two domains identified as the minor domains in each cycle. The major domain dominates two-thirds of the testing time. This aspect of PISA design provides a thorough analysis of achievement in each domain every nine years and a trend analysis every three years (OECD, 1999b, 2004b, 2006). From the research design perspective, PISA is a cross-sectional survey with all 15-year-old students in participating countries as the desired population. It samples the student participants by using multi-stage probability sampling.

2.2.3.4 The PISA Mathematical Literacy Framework

The OECD (2006) defines PISA mathematical literacy as

> an individual’s capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgements and to use and engage with mathematics in ways that meet the needs of that individual’s life as a constructive, concerned, and reflective citizen. (p. 72)

In accordance with its definition of mathematical literacy, PISA designs the assessment items by considering their relevance to three elements: situations or contexts, mathematical content, and competencies (OECD, 1999b, p. 41; 2004b, p. 24; 2006, p. 72). Table 2.3 summarizes the elements, attributes, and objectives of the PISA mathematical literacy framework (OECD, 1999b, p. 41; 2004b, p. 24; 2006, p. 72).

Table 2. 3. Structural Features of the PISA Mathematical Literacy Framework

Elements	Attributes	Objectives
Situations or Contexts	• Personal • Educational/Occupational • Public • Scientific	To assess students' ability to approach the items set in different contexts of the real-world situations
Mathematical Content	• Space and shape • Change and relationship • Quantity • Uncertainty	To assess students' mathematics content knowledge and skills
Competencies	• Thinking and reasoning • Argumentation • Communication • Modelling • Problem posing and solving • Representation • Using symbolic, formal and technical operation • Use of aids and tools	To assess students' mathematical competencies objectively

Because PISA is designed to assess the knowledge and skills required of 15-year-old students in their future, the PISA mathematical competencies call for students to demonstrate their mastery of mathematical processes and knowledge as well as their ability to apply them in solving problems in unfamiliar situations. Further, a notable feature in the PISA mathematical literacy framework is its inclusion of Situations or Contexts. The Situations or Contexts element offers a unique way to simulate a real-life situation for designing the assessment items. This also makes it possible for PISA to incorporate the contents and competencies in its assessment items in accordance with these situations. This aspect of PISA is either vaguely implicit or not available in NAEP and TIMSS.

Further, PISA described the cognitive activities encompassed in the eight competencies in terms of three competency clusters: the reproduction cluster, the connections cluster, and the reflection cluster (OECD, 2006). The description was based on the kinds of cognitive demands required to perform the cognitive activities contained in a mathematical problem. The reproduction cluster is the lowest level of the cognitive demand, consisting of standard representation and definitions, routine computations, routine procedures, and routine problem solving. The connections cluster is the intermediary level of the cognitive demand, involving modelling, standard problem solving, translation and interpretation, and multiple well-defined methods. The reflection cluster is the highest level of the cognitive demand, and it has complex problem solving and posing, reflection and insight, original mathematical approach, multiple complex methods, and generalization as the key features of cognitive demand.

2.2.3.5 PISA and Its Impacts and Critical Responses

To understand how successful PISA has been, it is desirable to review some responses to the PISA reports by some of its participant countries.

In Germany, PISA 2000 stirred unrest among German educators, politicians, the general public, and schools by what was known as the PISA shock, comparable to Sputnik shock, the French Revolution, and *A Nation at Risk* (Ertl, 2006, cf. Ostermann, 2005; Herrmann, 2004 and Gruber, 2006). The PISA shock resulted in immediate debates on the state of the German education system that eventually culminated into major reform initiatives (Ertl, 2006; Lange, 2002; McGaw, 2008b; OECD, 2004d).

In the U.K., the first report of the PISA 2000 surprised the stakeholders of its education system when the report ranked the performance of its 15-year-old students ahead of Switzerland's 15-year-old students (Prais, 2003). Prais (2003) cited the lack of connection between PISA frameworks and school curriculum, the use of age by PISA as the criterion for selecting its participants, the PISA's disregard of differential maturity rates of the participants, and the low response rates from the participants as the reasons for the unexpected performance by the U.K.'s 15-year-old students. Prais's critiques of PISA 2000 prompted an equally strong rejoinder by Adams (2003). After making line-by-line responses to Prais's observations, Adams (2003) concluded that Prais's criticisms were empirically not justified and largely attributable to his incomplete understanding and knowledge of the methodology of international studies such as PISA. However, the views exchanged between Prais and Adams reveal that PISA is vulnerable to adverse criticism despite its technical validity.

In Greece, students' low attainment level in PISA prompted Hatzinikita, Dimopoulos, and Christidou (2008) to study the alignment between the PISA science test items and the Greek school science textbooks. Hatzinikita et al. (2008) concluded that the textual materials employed in PISA and in the Greek school textbooks were oppositional in nature, and highlighted this as the potential factor for Greek students' low attainment levels in PISA.

In Australia, PISA's definition of scientific literacy was thought to be too ambitious for the 15-year-old students, especially when they have not learnt various aspects in their school curriculum (Fensham, 2002). However, the encouraging success of the students in PISA science in spite of this initial reservation fostered new debates amongst Australian educators. Fensham (2002) identified the following areas for positive implications of PISA on the Australian education system: aims for school science; curriculum structure; curriculum content; pedagogy; interest in science; assessment of science learning; SES and science learning; affective responses to the natural world; accessing scientific information and science and technology.

Follow-up studies of the PISA students in Australia, Canada, and Denmark revealed that the PISA reading performance of 15-year-old students is a very strong predictor for a successful transition to higher education at age 19 (OECD, 2007a), indicating the capability of PISA to predict the preparedness of 15-year-old students for successful participation in life beyond school contexts. A national educational assessment model that this study seeks to develop for Bhutan aims at a similar impact.

In summary, it is apparent that PISA has stimulated changes at various levels of the education systems of participating countries, indicating PISA's ability to drive educational reforms as the countries prepare their youth for the twenty-first century. The changes are most visible in areas such as teaching, learning, policies, curriculum, and resource allocation. This shows that PISA is a well-conceptualized programme to evaluate a nation's education system with reference to the education systems of other participating countries. A national educational assessment model that this study seeks to develop should have similar significance for Bhutan. Thus, PISA makes itself an ideal model for this new national educational assessment model.

2.3 Comparison of the Design Elements of NAEP, TIMSS, and PISA

NAEP, TIMSS, and PISA have a common goal of providing relevant educational stakeholders with research-based knowledge to guide educational policy, decisions, and interventions. Some of the factors that helped NAEP, TIMSS, and PISA succeed in their goals and objectives are their designs. This section presents the similarities and the differences among NAEP, TIMSS, and PISA.

2.3.1 Similarities and Differences in the Design Elements of NAEP, TIMSS, and PISA

TIMSS was first implemented in 1959, NAEP in 1969, and PISA in 1997. In passing, it might be noted that it would be reasonable to expect some precedence effects in NAEP from TIMSS and in PISA from both TIMSS and NAEP. The presence of precedence effects may also affect the reliability of the programmes. Table 2.4 depicts the similarities and the differences among NAEP, TIMSS, and PISA.

Table 2. 4. Comparison of Design Elements of NAEP, TIMSS, and PISA Frameworks

Design/Implementation Element	NAEP	TIMSS	PISA
Research type	Cross-sectional survey	Cross-sectional survey	Cross-sectional survey
Student Sampling	Random sample	Random sample	Random sample
Item Sampling	Matrix samples	Matrix samples	Matrix samples
Item Types	Multiple choice and constructed response	Multiple choice and constructed response	Multiple choice and constructed response
Test length	1.5 to 2 hours	1 to 1.5 hours	2 hours
Scaling	IRT-three parameter model	IRT-one parameter model	IRT- one parameter model and partial credit model
Alignment with curriculum	Assesses intended, implemented, and achieved curriculum	Assesses intended, implemented, and achieved curriculum with most focus on achieved curriculum	Assesses intended curriculum in terms of students' preparedness for life
Accommodation for special needs students	Yes	Exclusion upto 10% for special needs students allowed	Exclusion upto 2% allowed
Adaptation for limited English proficient students	Bilingual	30 languages + English and 11 cultural adaptations for English	Exclusion upto 2% allowed
Achievement levels	Three: basic, proficient, advanced	No, uses percentiles (5th, 25th, 50th, 75th, 95th) to describe the levels	Six: called levels 1to 6

Target population	Grades 4, 8 and 12	Grades 4, 8 and 12	15-year-olds in any grade
Assessment instruments	Subject papers; teacher, student and, school questionnaires	Subject papers; teacher, student school, and curriculum questionnaires	Subject papers; student, school, ICT, and parent questionnaires
Assessment cycle	Repeat every second year	Repeat every fourth year	Repeat every third year
Scope	National	International	International

From Table 2.4, it is clear that TIMSS, NAEP, and PISA are cross-sectional surveys, and that they use similar sampling designs and item formats. The existence of such similarities in TIMSS, NAEP, and PISA indicates a sustained precedence effect, confirming the validity of the research design, sampling methods, and item formats used by TIMSS, NAEP, and PISA.

Table 2.4 also shows that TIMSS and NAEP are similar in terms of their alignment with the national school curricula of participating countries and states. However, PISA differs widely from both TIMSS and NAEP in its alignment with the national school curricula of participating countries. This difference is largely due to PISA's research questions stated elsewhere in this chapter. Further, PISA does not use a teacher questionnaire unlike TIMSS and NAEP. The reason for this is that PISA is not completely aligned to the national curricula, thereby making it difficult for PISA to attribute student achievement to teacher quality (OECD, 2009a). However, there are ways to assess the alignment between an assessment programme such as PISA and a national curriculum, as demonstrated in this study in Chapter 5.

2.4 General Views on International Large-Scale Assessments

Researchers differ in their views about international large-scale assessments as much as countries differ in their reasons for participating in such assessments. Their views and rationales have the potential to guide the countries who wish to participate in one of the existing international large-scale assessments or who want to design a similar assessment programme. Therefore, this section examines critically the debates on international large-scale assessments and the reasons expressed by some countries for participating in one or more international large-scale assessments, with the view to using the debates to guide the development of a national educational assessment model for Bhutan.

2.4.1 Debates on Large-Scale Assessments

Researchers differ markedly in their advocacy or criticisms of the use of international large-scale assessments. The differences are visible in areas such as the population definition, the gap between school curricula and domains of assessment, and the translation of language. Researchers also caution about the use of correlational information for explaining cause and effect relationships. Furthermore, incorporating different assessment domains in a single assessment paper is identified as a potential source of bias. The following paragraphs examine these issues further.

Researchers caution against using correlational information generated from large-scale assessment data for causal interpretation (Taylor, 2000; Wenglinsky, 2002). Cross-sectional design of international large-scale assessments enables correlational analyses, but real explanatory power can only come from longitudinal investigations (Taylor, 2000; Wenglinsky, 2002). Therefore, cause and effect relationships from large-scale assessments like TIMSS, NAEP, and PISA is beyond their scope. Nevertheless, the descriptions of the educational circumstances of students at various achievement levels provided by TIMSS, NAEP, and PISA, when considered in the light of research from other sources, are known to have guided public discussions and policy actions (Harlen, 2001; McGaw, 2008b).

NAEP and TIMSS sample students by grade (IEA, 2008; NAGB, 2001), while PISA samples students by age (OECD, 2002, 2005d, 2009b), leading to debates among researchers (Egelund, 2008; Jenkins, 2000; Kitchen, 2000; McGaw, 2008a; Michael, 2001; Taylor, Jenkin, Curry, & Swennerton, 2000; Wiley & Wolfe, 1992). Grade-based sampling of students presumes some uniformity of formal schooling across countries and allows the study of teacher and school effects, but the way grades are allocated in schools differs from country to country. Age-based sampling of students seems better for comparability, but children begin school at different ages in different countries. Adams (2003) makes a persuasive point that age-based sampling enables PISA to assess the yield of education at an age that is common across countries. However, Wagemaker (2008) convincingly defends the use of grade-based sampling by TIMSS because it allows researchers to measure a fixed-period of learning across education systems. Wagemaker (2008) demonstrates empirically that age and student achievement do not correlate and that the performance of 15-year-old students depends on their grade. Wagemaker (2008) also claims that testing grade-based target populations allows the direct assessment of students' curriculum-based knowledge and skills and the analyses of the relation between student and teacher characteristics. McGaw (2008a) questions Wagemaker's claim because TIMSS assesses students' knowledge and skills acquired over four or eight years of schooling, during which it is likely that students have been taught by different teachers, making it difficult to attribute student achievement to any particular teacher.

The debates on the use of students' age or grade as criterion for sampling relate to the assessment goals. McGaw (2008b) notes that grade-based sampling reflects the hierarchical aspects of school organizational structures, while sampling by age includes all students at the sample age regardless of their grade levels. This implies a special significance for assessment goals in deciding on a choice of age- or grade-based sampling. Egelund (2008) convincingly notes that age-based sampling is preferable for international comparisons at the end of compulsory education, while grade-based sampling is favourable for assessment programmes that seek information about curricular content and policy strategies of national systems. Michael (2001) notes that age and grade have different effects across different countries and suggests the use of statistical procedures to adjust the outcomes of international tests for age and grade.

The challenges associated with translation of English to other languages for large-scale assessments are emphasized as threats to test validity (Michael, 2001). Jenkins (2000) states that the processes involved in translation are difficult, complex, and subtle. Nardi (2008) and Wiliam (2008) stated that the use of English as the original language of PISA items and its translation into other languages possibly introduces the elements of cultural bias in PISA. Taylor (2000), noting the structural and semantic difference among languages, suggested the use of bilinguals and back translation to solve the problems associated with translation. McGaw (2008a), referring to the methods of translation used in PISA, convincingly clarifies that translation should focus on establishing task consistency between source language and target language rather than focussing on language consistency between the two.

Incongruity between school curricula and assessment domains is seen as a barrier to using the findings from the latter to improve the former (Dohn, 2007; Jenkins, 2000; Nardi, 2008; Orpwood, 2000; Prais, 2003). It has been recommended that international comparative assessments incorporate longitudinal design elements and curriculum-based test items (Goldstein, 2004; Goldstein, Bonnet, & Rocher, 2007; Goldstein & Thomas, 2008). Noting the importance of congruity between large-scale assessments and school curricula, Webb (1997, 1999, 2006) presents a convincing quantitative method of evaluating the congruence between large-scale assessments and school curriculum.

Dohn (2007) states that the method chosen by PISA does not constitute an adequate operationalisation of the question of inquiry, indicating that PISA is not able to assess students' preparedness for life beyond school contexts. Dohn (2007) states, "contrary to the claims of PISA, PISA is not an assessment of the knowledge and skills for life of students, but only of knowledge and skills in assessment situations" (p. 1). Dohn's observation is reasonable but PISA attempts to address this issue by providing simulated real-world situations for its test items, which are as close as possible that a test item can come in replicating real-world situations, other than authentic assessments (Egelund, 2008).

Lie (2005) highlights the risk of adopting the frameworks of large-scale assessments as the mainstream curriculum in the schools; a phenomenon described as the washback effect (Hamp-Lyons, 1997). The OECD (2007a) also reminds researchers of the need to interpret and frame any influences from international large-scale assessments in national contexts. To a certain extent, this issue is a policy matter for each participating country. However, if a country decides to reform its education system due to a perceived deficiency, as revealed by the findings from a large-scale assessment, it is possible that an informed-policy decision, fully debated, can lead to a useful change in the national school curriculum.

The presentation of test items from different subjects that assess different skills in a single test paper by international large-scale assessments like TIMSS and PISA is seen as a threat to their validity (Jenkins, 2000). Jenkins (2000) states that testing student performance in science and mathematics in the same test requires students to switch their thinking between mathematics and science and between very different types of skills. This is likely to present unsolicited challenges to the test-takers, which may affect test validity. The effects of mathematics on science and vice-versa is hard to remove (Kitchen, 2000; Taylor, et al., 2000). A straightforward remedy to contain this issue is to have different papers for different subjects, assuming costs are not a consideration.

Because international achievement surveys are based on samples, their data are susceptible to both sampling and non-sampling errors, which affect the accuracy of the results. In their review of the IEA studies, Medrich and Griffith (1992) identify the response rates, the comparability of samples, and the non-sampling errors as three data-quality issues that have significant implications for data analysis. These issues are discussed in later chapters.

2.4.2 Reasons for Participating in Large-Scale Assessments

Countries participate in international large-scale studies for various reasons. This section examines critically why some countries participate in international large-scale assessments.

Medrich and Griffith (1992) provide a number of reasons for the United States's interest and participation in large-scale assessments. These are: the use of the international student achievement databases to study economic competitiveness; the widespread media coverage received by international student achievement studies; the interests in national and international educational perspectives, and the contextual details that accompany international large-scale assessments. Leimu (1992) notes that Finland participates in TIMSS to meet its political and research interests in its education system. Finland's preparedness for TIMSS and its interest in technical and management aspects of TIMSS are also pointed out as other reasons (Leimu, 1992). Bathory (1992) states that Hungary participates in TIMSS because it is able to train its researchers in advanced research methodologies and to develop close relationships with education in the West. In addition, Bathory (1992) states that Hungary's participation in TIMSS has fostered a culture of systemic evaluation of the Hungarian education system. Hussein (1992) refers to Kuwait's interest in using the comparative findings from international large-scale assessments to benchmark its science and mathematics curriculum and educational goals with those of other countries as the reason for its participation in TIMSS. Hussein (1992) also notes that information from TIMSS has guided Kuwait's initiatives in school mathematics reform. Watanabe (1992) identifies some reasons for Japanese participation in TIMSS as the interest in TIMSS design elements, the aim of developing national-level research capacity, the use of cross-national information to guide curriculum revision and teacher improvement programmes, and the opportunity for Japanese researchers to conduct secondary research with TIMSS data.

Summarising Section 2.4, general views on large-scale assessments suggest that the assessments are reliable and valid within the scope of their goals. The capability of the assessments to generate reliable and valid educational statistics is also evident in the reasons cited by countries for their participation in NAEP, TIMSS, and PISA.

2.5 Relationship Between Large-Scale, High-Stakes Tests and Large-Scale, Low-Stakes Assessments

Research has shown some correlations between large-scale, high-stakes tests and large-scale, low-stakes assessments (Cizek, 2001; Phelps, 2005). This allows researchers to use the merits and the demerits of large-scale, high-stakes tests to improve large-scale, low-stakes assessments.

A significant positive relationship between the scores on high-stakes test and the scores on large-scale assessments like TIMSS and NAEP is reported by Cizek (2001) and Phelps (2005). It is shown that the students in countries and states that require students to pass curriculum-based external exit examinations in order to graduate learn more than their peers who do not take such examinations. The research findings are summarised in Table 2.5 (Phelps, 2005, p. 78).

Table 2. 5. Learning Gains in Large-Scale, Low-Stakes Tests Due to Exit Examinations

Large-Scale Assessments	Grade-Level Equivalent Gains for Students who took Exit Exams
NAEP, Mathematics (New York and North Carolina as compared with other states, 1998)	0.4

NAEP, Science (New York and North Carolina as compared with other states, 1998)	0.5
NAEP, Reading (New York and North Carolina as compared with other states, 1998)	0.7
TIMSS, Mathematics (40 nations, 1995)	1.0
TIMSS, Science (40 nations, 1995)	1.3

These findings suggest that large-scale low-stakes assessments like TIMSS, NAEP, and PISA are reliable and valid measures of students' knowledge and learning competencies, and that they have a similar bearing on education systems to those of large-scale, high-stakes tests. However, high stakes tests have often been criticised by researchers and these critiques will now be reviewed.

2.5.1 Critiques of Large-Scale, High-Stakes Tests

Various researchers have noted both positive and negative effects of high-stakes tests. Some of their observations (Cizek, 2005; Goodman & Hambleton, 2005) are summarized in Table 2.6.

Table 2. 6. Positive and Negative Effects of High-Stakes Tests

Positive Effects	Negative Effects
Tests bring about focussed professional development	Makes students despair, cry, vomit, abandon studies and devalue grades
Develop inclusive tests	Causes frustration and anxiety in students
Educate teachers on testing technologies and their use in teaching and learning	Diminishes student's self-esteem and fosters negative attitudes to tested content
Develop students' performance database for identifying bases of success	Tests fail to measure what is important to K-12 education
Provide access to informed educational options for parents and students	Place too much emphasis on a single test score
Nurture accountability systems	Tests used as the only educational accountability system
Foster intimacy between educators and their disciplines	Tests set standards which are too high
Enhance quality of tests by funding psychometric research and increased student learning	Tests are full of biased items

On close examination of Table 2.6, it is clear that the positive effects of high-stakes tests are associated with educational improvements at student, teacher, school, and systemic levels. On the other hand, the negative effects are concerned more with students' readiness for the tests and their use in establishing accountability in the system. The negative effects of high-stakes tests indicate that the candidates of international large-scale assessments should not be subjected to decisions such as pass and fail. Further, the negative effects of high-stakes tests show that the findings from the international large-scale assessments should not be used in addressing accountability in participant schools. At this point, an inference is that the international large-scale assessments should be kept as low-stakes assessments for both assessment candidates and participant schools.

Phelps (2005) cites the following positive conclusions from different meta-analyses and research syntheses on high-stakes testing and its consequences on students' achievement:

- clear performance targets and goal-setting to reach them substantially increase productivity (cf. Locke & Latham, 2002);
- achievement gains are almost always higher with testing than without; the optimal amount is more than weekly (cf. Bangert-Drowns, Kulik & Kulik, 1991);
- feedback from tests improves achievement substantially, not only by identifying and clarifying weaknesses but also in disabusing those students who are complacent due to overconfidence (cf. Kulik & Kulik, 1998);
- mastery learning, and the testing that is an essential part of it, produces substantial achievement gains (cf. Kulik & Kulik, 1987);
- mastery learning produces substantial achievement gains (cf. Guskey & Gates, 1986);
- in general, higher standards lead to greater effort, in part because students tend not to take seriously work that adults do not seem to take seriously. However, there are limits; set the standards too high and some students may not try (cf. Natriello & Dornbusch, 1984);
- the more intense the experience, the more rapid the learning, and testing helps to intensify the experience (cf. Carroll, 1955); and
- studies showed that systematic reporting of test results assisted students (ninth grade) in developing greater understanding of their interests, aptitudes, and achievements (cf. Kirkland, 1971). (pp. 72-76)

This summary shows that student achievement is related to factors such as mastery learning, performance targets and goal setting, use of feedback, and learning experiences. Further, and more importantly, the summary shows that the use of cognitive and non-cognitive variables in large-scale assessments facilitate researchers in understanding the relation between student achievement and the factors that are known to influence student learning.

Stecher (2002, pp. 86-87), as shown in Table 2.7, synthesizes both positive and negative effects of large-scale, high-stakes testing at four levels: students; teachers; administrators; and policy makers.

Table 2. 7. Effects of High-Stakes Testing on Students, Teachers, and Administrators

Positive Effects	Negative Effects
Effects on Students	
Provide students with better information about their own knowledge and skills Motivate students to work harder in school Send clearer signals to students about what to study Help students associate personal effort with rewards	Frustrate students and discourage them from trying Make students more competitive Cause students to devalue grades and school assessments
Effects on Teachers	
Support better diagnosis of individual student needs Help teachers identify areas of strength and weakness in their delivery of curriculum Help teachers identify content not mastered by students and redirect instruction Motivate teachers to work harder and smarter Lead teachers to align instruction with standards Encourage teachers to participate in professional development to improve instruction	Encourage teachers to focus more on specific test content than on curriculum standards Lead teachers to engage in inappropriate test preparation Devalue teachers' sense of professional worth Entice teachers to cheat when preparing or administering tests

Effects on Administration	
Cause administrators to examine school policies related to curriculum and instruction Help administrators judge the quality of their programmes Lead administrators to change school policies to improve curriculum or instruction Help administrators make better resource allocation decisions, e.g., provide professional development	Lead administrators to enact policies to increase test scores but not necessarily increase learning Cause administrators to reallocate resources to tested subjects at the expense of other subjects Lead administrators to waste resources on test preparation Distract administrators from other school needs and problems
Effects on Policy Makers	
Help policymakers to judge the effectiveness of educational policies Improve policymakers' ability to monitor school system performance Foster better allocation of state education resources	Provide misleading information that leads policymakers to suboptimal decisions Foster a "blame the victims" spirit among policymakers Encourage a simplistic view of education and its goal

The positive effects of high-stakes tests, as highlighted in Table 2.7, have potential to assist in setting goals and developing research questions for international large-scale assessments. The underlying cause for the negative effects of high-stakes tests across all four levels is the use of test results to address accountability. Therefore, to avoid such negative effects, it is imperative that large-scale assessments are perceived as low-stakes, high-value assessments by students, teachers, administrators, and policy-makers; with low-stakes maintained by dissociating accountability from assessment for learning, and high-value exemplified by using assessments to guide educational decision-making and the development of teaching and learning interventions.

2.6 Putting together the Notable Design Principles of Large-Scale Assessments

Drawing on the earlier reviews, this section combines the notable framework principles and design elements of NAEP, TIMSS, and PISA, with the view to using them in developing a national educational assessment model for Bhutan.

2.6.1 Notable Design Principles for Bhutan

The NAEP, the TIMSS, and the PISA frameworks have many common elements such as the consensual approach, the hierarchies of expertise, the assessment content domains, the accommodation of cultural contexts, and the option of using students' age or grade as a central criterion for sampling.

NAEP, TIMSS, and PISA have similar organizational structures that are formed with various expert committees and representatives of different stakeholders. Representatives of stakeholders ensure the inclusion of broad perspectives in the assessment programmes, while expert committees develop assessment frameworks and write test items in line with the stakeholders' perspectives. The involvement of representatives of stakeholders and expert committees in assessments develops wider ownership and underpins the validity of the assessments, feature that is desirable for a national educational assessment model that this study seeks to develop for Bhutan.

A contentious debate among researchers in large-scale assessments is the use of students' age or grade as the basis for defining the population at which the assessments are targeted. This study will deal with the prospective student population in terms of their grade, not their age. A grade-based sampling of students makes it possible to link data from large-scale assessments to school curriculum and other factors related to school effectiveness, which is one of the objectives of this study.

NAEP, TIMSS, and PISA have assessment frameworks that are similar in their purposes and definitions to a conventional test blueprint, describing item formats, item difficulty levels, and item contents. This study will develop a similar test blueprint.

NAEP, TIMSS, and PISA use achievement levels to describe students' subject knowledge and skills. Because PISA claims to assess the yield of education at the end of compulsory education, which occurs when children are 15-year-olds in most countries, the achievement levels described by PISA are preferred for use in this study. PISA also explicitly claims to assess the preparedness of students for life-long education and its capacity to assess students' higher-order thinking skills is reported to be greater than that of TIMSS. These features of PISA fit well with the educational policy priorities of Bhutan that were outlined in Chapter 1.

NAEP, TIMSS, and PISA all use multi-stage probability sampling methods as the preferred sampling design. This sampling design allows researchers to analyse data from large-scale assessments at different levels of aggregation and enhances sampling efficiency. A similar sampling design will be used in this study.

The critiques of high-stakes tests underline the importance of keeping the large-scale assessments as low-stakes assessments to avoid the concerns associated with high-stakes tests. NAEP, TIMSS, and PISA are all maintained as low-stakes, large-scale assessments. Therefore, a national educational assessment model that this study seeks to develop for Bhutan will be strategically profiled as a low-stakes, high-value assessment model with the capacity to guide educational policy-decisions and teaching and learning interventions.

2.7 Chapter Summary

A review of NAEP, TIMSS, and PISA revealed that large-scale assessments have been instrumental in guiding educational policy decisions, driving educational reforms, and facilitating educational debates across the world. Such instrumental roles of NAEP, TIMSS, and PISA are evident in their objectives and reports. A factor that enabled NAEP, TIMSS, and PISA to succeed in their objectives is their design.

Further, consensual approaches to design by involving experts and stakeholders, assuming low-stakes, high-value status, incorporating both cognitive and non-cognitive aspects of teaching and learning, using achievement levels to profile students' knowledge and skills are consistently used in NAEP, TIMSS, and PISA.

Despite their immense contributions to educational reforms, large-scale assessments do have limitations. The cross-sectional survey design aspects of NAEP, TIMSS, and PISA restrict the researchers from interpreting the findings from these large-scale assessments to correlational interpretations, not causal interpretations. In addition, the use of multi-stage sampling design by NAEP, TIMSS, and PISA makes it impossible to generate an individual student report card. Inconsistencies between the content of large-scale assessments and school curriculum also restrict the findings from being used to provide feedback on school curricula.

However, researchers have proposed relevant measures to counter the limitations of large-scale assessments, with the measures ranging from defining assessment purposes to survey design. Researchers have proposed longitudinal studies to enable causal interpretations from large-scale assessments. For the findings from large-scale assessments to be relevant to school curriculum, researchers recommend grade-based sampling rather than age-based sampling. Obviously, for an individual student report card, a census would need to be used in place of a sample survey.

The PISA-like approach has been proposed as the model for developing a national educational assessment model for Bhutan. A PISA-like approach is preferable to direct replication of PISA because some amendments would be necessary to make it suitable for Bhutan.

2.8 Chapter Conclusion

This chapter has examined many aspects of large-scale assessments. Large-scale assessments are developed consensually by involving stakeholders and expert committees, with the former providing goals and objectives of assessments to the latter. Large scale-assessments use frameworks as a platform for specifying their goals, contents, and test items. The goals of PISA are more compatible with the educational policy priorities of Bhutan than the research questions of NAEP and TIMSS. Therefore, a PISA-like approach will guide a national educational assessment model that this study seeks to develop for Bhutan.

Large-scale assessments play a pivotal role in guiding educational policy decisions and driving educational reform. However, large-scale assessments have limitations. Large-scale assessments are cross-sectional surveys. Therefore, large-scale assessments do not provide causal interpretations. In addition, it is not possible to generate individual student report cards from data collected through sample-based large-scale assessments. Resources permitting, this limitation of large-scale assessments can be addressed by using census surveys, instead of using sample surveys. The irrelevance of large-scale assessments to school curricula also limits the findings from large-scale assessments for use in guiding school curriculum reforms. A solution to this problem is to evaluate the alignment between school subject standards and test-items of large-scale assessments using a set of alignment criteria. The use of large-scale assessments to determine accountability in schools, results in maladaptive behaviours by the parties affected by the findings from the assessments. A solution to this problem is to profile large-scale assessments strategically as low-stakes, high-value assessments, with the aim of gaining willing participation from students, teachers, and schools.

All these aspects of large-scale assessments have significant potential to guide the development of a national educational assessment model that this study seeks to develop for Bhutan.

Chapter 3

CONTEXTUAL VARIABLES

3.1 Introduction

This chapter examines critically the different traditions of educational effectiveness research, and reviews the contextual variables in the literature on educational effectiveness, to assist in developing a system-wide educational assessment model for Bhutan.

The first section of the chapter examines the different strands of educational effectiveness research, with the aim of developing a conceptual national educational assessment model for Bhutan. The second section identifies contextual variables used in NAEP, TIMSS, and PISA to guide the Bhutanese review of its various contextual variables. Guided by the contextual variables identified in the second section, the third section reviews critically the various contextual variables in the literature on educational effectiveness, with the objective of relating them to student achievement in mathematics. The fourth section of the chapter presents a national educational assessment model for Bhutan.

3.2 Review of Educational Effectiveness Research

Research studies that deal with the factors that promote educational success are broadly referred to as educational effectiveness studies. Depending on the disciplinary backgrounds and the areas of interests, researchers developed different models of educational effectiveness studies to suit their research goals and objectives. Researchers in educational effectiveness have used the three disciplines of economics, sociology, and psychology to explain why a student, or a teacher, or a school, or a country performs better in education than another when their background characteristics are adjusted (Creemers & Kyriakides, 2008; Scheerens, 1992). Economists emphasise productivity as the main criterion of effectiveness of the optimal relationship between resource inputs and output values, sociologists emphasise equity and organizational theories as the main attributes of effectiveness, and psychologists use theories of teaching and learning to characterize educational effectiveness. Such diverse perspectives on educational effectiveness indicate that educational effectiveness research is a complex phenomenon that needs a trans-disciplinary eclectic approach.

3.2.1 Concept of Educational Effectiveness

Researchers in educational effectiveness conceptualize effectiveness from a range of disciplinary perspectives in relation to student outcomes. Reynolds et al. (1994) and Creemers and Kyriakides (2008) delineate quality and equity as two dimensions of effectiveness, with quality implying the within-school difference and equity implying the between-school difference in student outcomes. According to these dimensions, an effective school will have a large positive difference in student achievement with respect to other schools and a very small difference in student achievement within the school (Creemers & Kyriakides, 2008). Creemers and Kyriakides (2008, p. 3) consistently suggest that educational effectiveness studies attempt to “establish and test theories that explain why and how some schools and teachers are more effective than others”.

Scheerens (1992, 1997, 2000) viewed school effectiveness from the vantage point of pluralistic and relativistic perspectives of organizational effectiveness. The pluralistic perspective ensues from the availability of different criteria to judge organizational effectiveness, with the criteria being productivity, adaptability, involvement, continuity, and responsiveness. From a pluralistic perspective, a school may be effective when it (a) produces high student output; (b) adapts to fluctuation in input resources; (c) displays highly motivated people evident in their involvement in school activities; (d) strives to establish formal structures; and (e) maintains an interdependent relationship and balance of power by networking with stakeholders (Scheerens, 1992, 1997, 2000). From a relativistic perspective, the effectiveness of a school may be judged solely in terms of its productivity by taking other criteria as antecedent conditions. Similar to the relativistic perspective is the contingency perspective which attributes the predominance of a particular effectiveness criterion to the influence of an organization’s goals, structures, life-history, and other contingencies (Scheerens, 1992). Scheerens (1992) conclusively notes that different criteria of organizational effectiveness can be ordered as a means to an effective end, with productivity as the ultimate criterion. Educational effectiveness, therefore, refers to the extent to which educational processes (means) lead to attainment of educational goals (productivity). Creemers and Kyriakides (2008) also conclude that all possible malleable features of the functioning of schools have to be identified and judged in terms of the school’s productivity. As varied as the attributes of effectiveness are, and as intensive as the researchers’ quest for the factors that enhance student outcomes is, the educational effectiveness research too has evolved over the years.

The educational effectiveness research evolved into various research traditions. Researchers (e.g., Creemers & Kyriakides, 2008; Reynolds, Teddlie, Creemers, Scheerens, & Townsend, 2000; Scheerens, 1992) have grouped educational effectiveness studies into the following traditions: equality of educational opportunity (e.g., Coleman et al., 1966; Jencks et al., 1972; Lynch & Baker, 2005; Wiggan, 2007); educational production functions (e.g., Greenwald, Hedges, & Laine, 1996; Hanushek, 1997; Hedges, Laine, & Greenwald, 1994; Lopez, 2007; Walberg, 1980, 1984; Walberg, Haertel, Pascarella, Junker, & Boulanger, 1981); instructional effectiveness (e.g., Fraser, Walberg, Welch, & Hattie, 1987; Harris, 1998; Marzano, 1998; Schroeder, Scott, Tolson, Huang, & Hsuan, 2007; 2007; Walberg, 1984; Wayne & Youngs, 2003; Wise & Okey, 1983); effective schools (e.g., Barber & Mourshed, 2007; Sammons, Hillman, & Mortimore, 1995); and school effectiveness (e.g., Creemers & Kyriakides, 2008; Reynolds, et al., 1994; Scheerens, 1992; Scheerens & Bosker, 1997; Teddlie & Reynolds, 2000). These research traditions are examined critically by focussing on their assumptions, empirical evidences, merits, and demerits.

3.2.2 Educational Opportunity Research

The Coleman Report (Coleman, et al., 1966), a large-scale survey that involved over 4,000 schools, 60,000 teachers, and 600,000 pupils in schools across the United States, laid the foundation of school effectiveness studies. The Coleman Report, based on an input-output model with input being mainly resources (e.g., school building, library) and output being student achievement, concluded that "schools bring little influence to bear on a child's achievement that is independent of his background and general social context" (Coleman, et al., 1966, p. 325). Given the controversial nature of the Coleman Report, Jencks et al. (1972) re-analysed the data collected by Coleman et al. (1966) and arrived at similar conclusions: schools did little to reduce the gap between rich and poor, or more able and less able students, and student achievement was primarily a function of student background.

However, apart from acknowledging students' SES, prior knowledge, and abilities as the correlates of student achievement, researchers disagree with the pessimistic view of the role of schools in improving student achievement (Scheerens, 1990). Scheerens (1990) attributes the pessimistic views of Coleman (1966) and Jencks (1972) to their use of very few schooling process variables as compared to their use of school resource variables. The controversial role of schools in improving student achievement has led to the emergence of other traditions of educational effectiveness research.

3.2.3 Educational Production Functions Research

Educational production functions research is similar to the equality of educational opportunity research in that both research traditions use input-output as the main basis for studying the effectiveness of schooling. However, the educational production functions research stresses the input-output relation in terms of costs of input resources and values of educational outputs, as commonly emphasized by economists. This salient feature of the educational productivity research tradition underscores its use of inputs measureable in monetary terms (e.g., teacher experience, student-teacher ratio, per pupil expenditure) (Scheerens, 1990, 1992). Also, the educational production functions research is based on the assumption that increased inputs should lead to increased outcomes (Creemers & Kyriakides, 2008).

Critiques of the educational production functions research note that the models used in it yield a disconcerting pattern of inconsistent and often insignificant results (Hanushek, 1979, 1986, 1997; Monk, 1992). It neither incorporates the nested nature of educational production (Hanushek, 1979, 1986; Monk, 1992) nor includes the dynamic nature of the educational production processes (Monk, 1992). Creemers and Kyriakides (2008) note that the difficulty of determining the monetary value of educational inputs and processes and the diverse educational outputs make purely economic approaches to educational analysis difficult and problematic. After reviewing 587 studies that used an educational production functions model, Hanushek (1989, 1997) concluded that a strong or systematic relationship between school expenditures and school performance is difficult to find. Researchers' discontent with educational production functions research led to the emergence of instructional effectiveness research.

3.2.4 Instructional Effectiveness Research

Instructional effectiveness research examines educational process variables (e.g., teaching behaviours) at the teacher or classroom level with reference to student outcomes (Scheerens, 1990). The use of the teacher or the classroom as the unit of analysis differentiates the instructional effectiveness research from the equality of educational opportunity research and the educational production functions research.

A number of instructional effectiveness research studies have shown a consistent and positive relationship between instructional variables and student outcomes (Harris, 1998; Marzano, 1998; Scheerens, 1992). Some of the consistent findings from instructional effectiveness research studies are: effective teaching is goal-directed; effective teaching consists of skills, strategies, and behaviours of teachers; an extensive repertoire of teaching models or strategies enhances teaching effectiveness; and teaching effectiveness is malleable (Harris, 1998; Marzano, 1998, 2001; Marzano, Gaddy, & Dean, 2000; Mortimore, 1993; Schroeder, et al., 2007; Seidel & Shavelson, 2007).

In summary, because instructional effectiveness research focuses on the teaching and learning, it offers a framework for conceptualizing educational process variables related to classroom, teacher, and student characteristics as correlates of student achievement.

3.2.5 Effective Schools Research

This research tradition involves the identification of schools that perform better than other schools in terms of student outcomes (Barber & Mourshed, 2007; Brookover, Beady, Flood, Schweitzer, & Wisenbaker, 1979; Marzano, 2001; Mortimore, Sammons, Stoll, Lewis, & Ecob, 1988; Scheerens, 1992). The schools identified as top performing schools are then further researched to find out salient features that might have enabled this outcome.

Scheerens (1990) notes that with the advent of the effective schools research, the unexplored ‘black box’ of what happens within schools is opened and school variables are revealed that include school organization, school culture, and school technology. In this way, the following school characteristics are consistently found to associate with effective schools: strong educational leadership; high expectations of pupil attainment; an emphasis on acquiring basic skills; a safe and orderly school climate; and the frequent evaluation of pupil progress (Creemers & Kyriakides, 2008; Scheerens, 1990, 1992). Although the effective schools research succeeded in highlighting the point that schools do matter or contribute to students’ cognitive, affective, and sensory motor skills development, the research had been a subject of academic discussion for some time.

Effective-schools research is criticised for using small samples and risking ‘Type I Error’ (i.e., observing a difference when in reality there is none) or ‘Type II Error’ (i.e., failing to observe a difference when in reality there is one) in identifying effective schools (Purkey & Smith, 1983). Effective schools studies aggregate data at the school level which may fail to show differential performances by different subgroups of students. This not only disregards the nested-layer feature of a school, but also implies that the effective schools research is not designed to assist schools in improving different subgroups of students (Purkey & Smith, 1983).

In summary, effective schools research enables comparison between schools, with the objective of improving ineffective schools by introducing the salient characteristics of effective schools. However, as noted by Purkey and Smith (1983), this research needed to improve its design. As a result, school effectiveness research has emerged that bears a close resemblance to effective schools research.

3.2.6 School Effectiveness Research

Scheerens (1992) defines the school effectiveness research tradition as the studies that seek to explain the differences in student outcomes in terms of specific school characteristics. Unlike the effective school research, school effectiveness research uses a random sampling design and a multi-level data analysis technique to include effective and ineffective schools and to account for the nested layers of school structures (Scheerens, 1992). In addition, in response to the critics concerned with the context-specificity and generalizability of the findings and recommendations originating from it, school effectiveness research emphasises school context as one of the important units of information about student outcomes (Scheerens, 1992). Scheerens’s Integrated Model of School Effectiveness is presented in Figure 3.1.

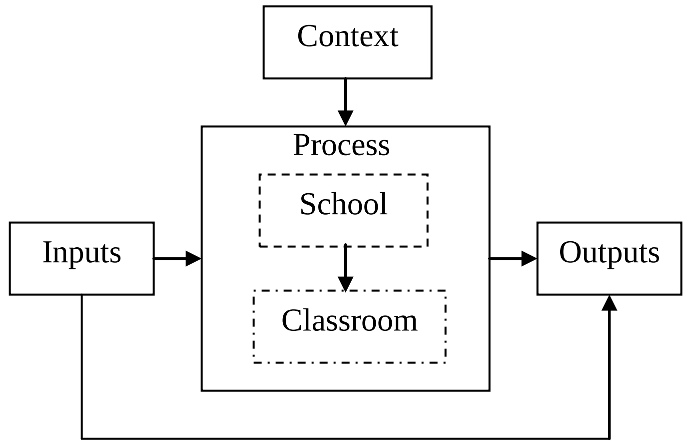


Figure 3. 1. The Integrated Model of School Effectiveness

As shown in Figure 3.1, the Integrated Model of School Effectiveness has the features of educational production functions and effective school research traditions, such as input-output, input-process, and output. This model also has multi-levels, such as classroom-, school-, and context-levels. The downward arrowheads in the model indicate that the higher levels facilitate conditions for the lower levels—an assumption that enables the study of cross-level interactions. The model also shows that output is determined by both input and process, while process is determined by input. Drawing on the empirical evidence from different traditions of educational effectiveness research studies, Scheerens (1990, 1997) fleshed out his Integrated Model of School Effectiveness with the variables in Table 3.1.

Table 3. 1. The Variables of the Integrated Model of School Effectiveness

Unit	Variables
Context	Achievement stimulants from higher administrative levels; development of educational consumerism; and co-variables like school size, student-body composition, school category, and urban/rural
Input	Teacher experience; per pupil expenditure; and parent support
Process	School Achievement-oriented policy; educational leadership; consensus, co-operative planning of teachers; quality of school curricula in terms of content covered, and formal structure; pressure for achievement; recruitment of qualified staff; financial and material characteristics of the school; orderly atmosphere; and evaluative potential Classroom Time on task; structured teaching; opportunity to learn; high expectations of pupils' progress; evaluation and monitoring of pupils' progress; and reinforcement
Output	Student achievement adjusted for previous achievement, intelligence, and SES

The Integrated Model of School Effectiveness accords with the necessary features of a school effectiveness model set out by Reynolds et al. (1994), except its capability to explain individual student gains. Reynolds et al. (1994) assert that a school effectiveness model must: explain individual student gains; aggregate upward; specify the relationships among variables contributing to those gains; be capable of being measured with reasonable accuracy over at least three points in time; and be capable of being analysed at appropriate levels. Exclusion of a specific level for students in the Integrated Model of School Effectiveness may subsume student characteristics at the classroom-level, implying that the model permits data analysis only at the classroom level, not at the student level of aggregation. The literature on educational effectiveness, however, abounds with claims that student characteristics (e.g., SES, motivation, and self-regulatory learning skills) not only affect student achievement, but also interact with the effectiveness factors operating at the classroom or school-level, implying that the Integrated Model of School Effectiveness needs reformulation.

Creemers (1994) developed a comprehensive model of educational effectiveness that includes student-level. Like Scheerens's model, Creemers's (1994) model has a multi-level structure, with context, school, classroom, but Creemers adds student as a fourth level. Creemers's (1994) comprehensive educational effectiveness model has four assumptions. First, time on task and opportunity used at the student-level are directly related to student achievement. Second, the quality of teaching, the curriculum, and the grouping procedures influence time on task and opportunity to learn. Third, teaching quality, time, and opportunity at the classroom-level are also influenced by factors at the school-level that may or may not promote these classroom factors. Fourth, student achievement is also determined by student factors such as aptitude, social background, and motivation.

Researchers have reported empirical evidence supporting the validity of Creemers's (1994) Comprehensive Educational Effectiveness Model, especially its multi-level nature and direct and indirect relationships between the levels and student outcomes (Creemers & Kyriakides, 2006, 2008; Kyriakides, 2008). However, Kyriakides (2008) pointed out that Creemers's (1994) Comprehensive Educational Effectiveness Model allows researchers to use different approaches to measure effectiveness factors that lead to inconsistent results with other models. Kyriakides (2008) noted that the main cause of inconsistency is that researchers describe effectiveness factors as uni-dimensional, instead of describing them as multi-dimensional. It was also observed that the studies conducted to test the validity of Creemers's (1994) Comprehensive Educational Effectiveness Model did not identify cross-level interactions between the factors at different levels (Kyriakides, 2008). Kyriakides (2008) strongly attributed the absence of cross-level interactions in the studies that tested Creemers's (1994) model to its oversight of the dynamic nature of effectiveness.

Using the multi-level nature of educational effectiveness and the existence of relationships between levels and outcomes as confirmed by the studies that tested Creemers's (1994) model as starting points, Creemers and Kyriakides (2008) developed a Dynamic Model of School Effectiveness. The model is shown in Figure 3.2.

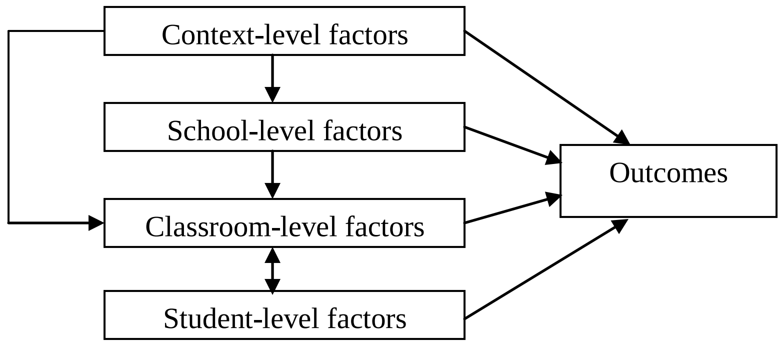


Figure 3. 2.The Dynamic Model of School Effectiveness

Creemers and Kyriakides (2008) incorporate the following assumptions in their Dynamic Model of School Effectiveness:

- the factors influencing student outcomes are multi-level;
- the influence of classroom-level factors on student outcomes is more direct and proximal than the influences from the factors operating at other levels;
- the higher levels provide an enabling environment for the lower levels;
- the relationship among school-level, context-level factors, and student outcomes develop and manifest themselves over time;
- the effectiveness factors operate differentially across levels in line with contingency theories;
- the relationship between some effectiveness factors and student outcomes are not always linear;
- within-level and between-level interaction effects exist among effectiveness factors.

Creemers and Kyriakides (2008) claim that, along with the identification of the educational effectiveness factors, an educational effectiveness model should explain various dimensions upon which the functioning of the factors can be measured, with the dimensions being frequency, focus, stage, quality, and differentiation aspects of effectiveness factors. Creemers and Kyriakides (2008) fleshed out Figure 3.2 with the variables shown in Table 3.2.

Table 3. 2. The Variables of the Dynamic Model of School Effectiveness

Unit	Factor
Context	National/regional policy for education; evaluation of policy; and the educational environment
School	School policy and evaluation of school policy
Classroom	Quality of teaching (orientation, structuring, modelling, application, questioning, assessment, management of time, classroom as a learning environment)
Student	Aptitude; perseverance; time on task; opportunity to learn; SES; gender; ethnicity; personality traits; expectations; thinking style; and subject motivation
Outcomes	Cognitive; affective; psychomotor; and new learning

Creemers and Kyriakides's (2008) Dynamic Model of School Effectiveness has all the three necessary characteristics of a school effectiveness model that are identified by Reynolds et al. (1994). However, the Dynamic Model of School Effectiveness does not have an input unit in it. It is clear from the literature on educational effectiveness that educational input resources, including teacher background characteristics, contribute to variance in student achievement even after adjusting for student background and prior learning (Greenwald, et al., 1996; Hanushek, 1997; Hedges, et al., 1994; Marzano, 2001; Scheerens, 2000; Wayne & Youngs, 2003; Wenglinsky, 2002). In addition, Ingvarson and Rowe (2007) presented a compelling case for conceptualising teacher quality in terms of teachers' subject-matter knowledge and pedagogical skills, in contradiction to Creemers and Kyriakides's (2008) conceptualisation of teaching quality only in terms of teachers' pedagogical skills. Therefore, without the input unit, Creemers and Kyriakides's (2008) Dynamic Model of School Effectiveness does not align with the literature on educational effectiveness.

3.2.7 A Skeletal National Educational Assessment Model for Bhutan

The strengths and weaknesses of the Integrated School Effectiveness Model and the Dynamic Model of School Effectiveness offer important guidelines for developing a national educational assessment model for Bhutan. First, the input-process-output paradigm of the Integrated Model of School Effectiveness is a case in point. Drawing on the literature of educational effectiveness research, Scheerens (1990, p. 62) sums up that a "context-input-process-output model is the best analytic scheme to systematize thinking on [educational] indicator systems". Therefore, a national educational assessment model should accommodate the discrete nature of educational productivity, with each production unit (input, process, and output) relating to various correlates of educational effectiveness. Second, a multi-level structure of a national education assessment model fits well with the conventional school organizational structures, and aligns well with the multi-level characteristic of the factors that affect student outcomes (Creemers & Kyriakides, 2008). Therefore, a national educational assessment model needs context-level, school-level, class-level, and student-level components, as in Creemers and Kyriakides's (2008) Dynamic Model of School Effectiveness.

While school-, class-, and student-levels are easily noticeable in a conventional school organizational structure, context-level influences are not always obvious. However, it is the context-specific educational goals, objectives, values, and their priority levels that define educational effectiveness. Reynolds (2006) notes that countries construe educational effectiveness factors differently. For instance, countries have different concepts of an effective instructional style as depicted in Table 3.3, adapted from Reynolds (2006, pp. 547-552).

Table 3. 3. Criteria of Instructional Effectiveness in Different Countries

Country	Effectiveness Criteria of Instructional Style
Canada	Collaborative; problem-based; teacher-directed
The United States	Classroom climate; class management; instructional delivery skills; use of innovative practices
Norway	Interactive teaching; group work; whole class involvement; social or academic focus; guidance
The U.K.	Proportion of whole class direct instruction; teacher management of groups; time use; interactive teaching; match of task/pupil
Ireland	Traditional/progressive; organized/disorganized
Taiwan	Inventive methods; curriculum match; interaction 'withitness'
The Netherlands	Number of discussed problems; whole class teaching; time use; feedback; assessment; high instructional time; efficient class management

As is clear from Table 3.3, countries differ in their views on educational effectiveness criteria. Relating Table 3.3 to Bhutan, a study in teachers' perceptions of classroom effectiveness, conducted by the Education Monitoring and Support Services Division (2003) of the Ministry of Education, reported that teachers viewed teaching competency (ability to plan and prepare lessons well), educational program (school curriculum), and classroom management as the key factors that influenced classroom effectiveness. The diverse views of teachers from different countries about instructional effectiveness underline the importance of including a specific context-level in a national educational assessment model. A unique context-level will enable the model to generate information about country-specific socio-cultural values, overall educational environment, and national education policy. As assumed by the Integrated Model of School Effectiveness and the Dynamic Model of School effectiveness (Creemers & Kyriakides, 2008), the multi-level nature of educational effectiveness is based on the knowledge that the conditions at higher levels facilitate the conditions at lower levels. These analyses suggest the following characteristics of an effective national educational assessment model: the model should be multi-level in nature; the model should be based on input-process-output paradigm; and the relationship between different levels of the model might be linear, curvilinear, or reciprocal. A skeletal national educational assessment model for Bhutan is thus presented in Figure 3.3.

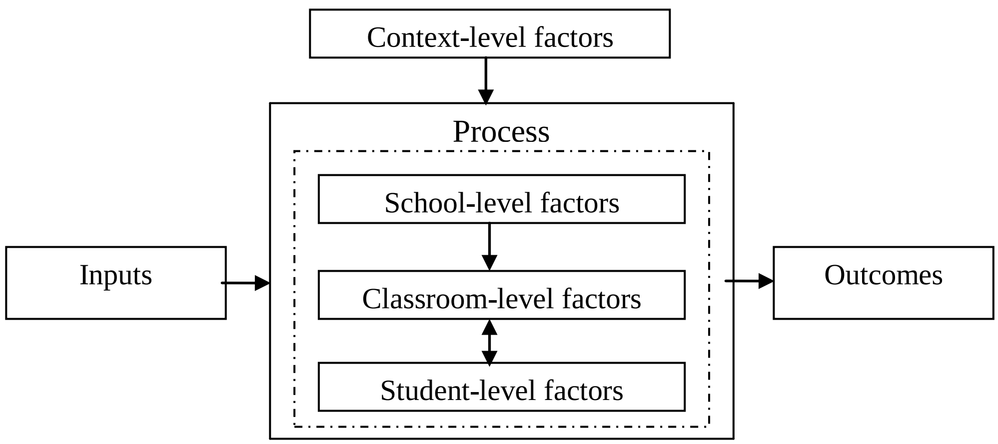


Figure 3. 3. National Educational Assessment Model for Bhutan

The proposed national educational assessment model in Figure 3.3 amalgamates Scheerens's (1992) Integrated Model of School Effectiveness and Creemers and Kyriakides's (2008) Dynamic Model of School Effectiveness. The proposed model is, therefore, based on the same assumptions formulated by Scheerens (1992) and Creemers and Kyriakides (2008) for their models. A comparison of the proposed national educational assessment model with the other two models is presented in Table 3.4.

Table 3. 4. Comparison of Models in Terms of Their Structures

Characteristics	Integrated School Effectiveness Model	Dynamic Model of School Effectiveness	National Education Assessment Model
Context-level	√	√	√
School-level	√	√	√
Classroom-level	√	√	√
Student-level	×	√	√
Input unit	√	×	√

Process unit	√	√	√
Output unit	√	√	√

As is shown in Table 3.4, the three models have considerable similarity. However, the models differ in certain areas. The Integrated School Effectiveness Model lacks the student-level, whereas the Dynamic Model of School Effectiveness lacks the input unit. By integrating the two models, the national educational assessment model bridges the gaps between these two models.

3.3 Contextual Variables of NAEP, TIMSS, and PISA

This section presents the main purposes for using certain contextual variables in large-scale assessments and an overview of criteria for selecting contextual variables. The section also presents a list of contextual variables that are used in NAEP, TIMSS, and PISA to guide the identification of contextual variables for this study.

3.3.1 Contextual Variables: Purposes and Criteria

Researchers have presented compelling reasons for using contextual variables in educational assessment programmes. Contextual variables assist in disaggregating student achievement by population sub-groups (Barton, 2002; NAGB, 2000). Contextual variables also support and help further research on factors and conditions that have been established by the research community as correlates of educational achievement (Barton, 2002; Grissmer, 2002; NAGB, 2000). Contextual variables assist in informing educational policy about the contexts of learning (Mullis, 2002; Walberg, 2002). Contextual variables help to evaluate the potential for bias in assessment results due to non-participation by students in assessment (Mullis, 2002; NAGB, 2003). The information obtained from contextual variables helps in tracking, over time, changes in contextual and instructional factors that are linked to student outcomes and guides the distribution of educational resources (NAGB, 2003).

Given their important purposes, researchers recommend specific criteria for incorporating contextual variables in educational assessment programmes. NAEP has set the following criteria for selecting contextual variables: relevance to the main purposes of NAEP; professional standards for reliability and validity; currency; broad public acceptability; and value in comprehending and explaining student achievement for improvement (NAGB, 2003). Contextual variables also must be based on the findings from the educational and social science research studies and meta-analyses (Barton, 2002; OECD, 2007a). Further, all contextual variables used in NAEP, TIMSS, and PISA are known to correlate with student outcomes (Mullis, et al., 2004; NAGB, 2006b; OECD, 2007a), which is another important criterion.

3.3.2 Contextual Variables used in NAEP, TIMSS, and PISA

NAEP, TIMSS, and PISA collect contextual information from a range of sources such as students, teachers, parents, curriculum specialists, and school principals. They demonstrated consistently that their contextual variables are correlated to student outcomes (Lee, Grigg, & Dion, 2007; Mullis, et al., 2005; Mullis, et al., 2004; NAGB, 2003, 2006b; OECD, 2006, 2007a). Table 3.5 lists the contextual variables used in NAEP, TIMSS, and PISA. The contextual variables are classified with reference to student, teacher, curriculum, and school.

Table 3. 5. Contextual Variables Used in NAEP, TIMSS, and PISA

NAEP Contextual Variables		
Student	Teacher	School
Socio-economic status	Race	School type
Race/ethnicity	Gender	Socio-economic status
Age	Experience	Grade structure
Gender	Credentials	Instructional days
Disability status	Academic qualification in teaching subject	Total enrolment
LEP	Professional qualification	Enrolment mobility
Enrolment status	Frequency of correction work	Grade retention
TV watching	Frequency of laboratory work	Graduation rates
Absenteeism		% of LEP
Language in home		% of students absent
Courses taken		% of teachers absent
Time spent on homework		Teacher retention
Self-efficacy		Post-secondary education
		Use of ICT

TIMSS Contextual Variables			
Student	Teacher	Curriculum Specialists	School

Home and school lives Classroom experiences Self-perception and attitudes Homework Out-of-school activities Use of ICT Home educational support Demographic information	Academic preparation and Certification Induction programme Professional development Demographic information Curriculum topics taught and time spent Instruction time Instructional activities Assessment and homework Use of ICT Use of calculator Emphasis on investigation	Formulation of curriculum Scope and content of the curriculum Organization of the curriculum Monitoring and evaluation of the implemented curriculum Curricular materials and support	School demographics School organization School goals Roles of the school principal Resources to support mathematics and science learning Technology, support, and equipment School social climate Parental involvement Teacher recruitment Teacher evaluation

PISA Contextual Variables		
Student	School	Parent [science related]
Background Learning and instruction Motivational factors Self-beliefs Value beliefs Subject related careers Use of ICT	School characteristics Admittance policies and instructional context School management School resources School activities	Students' past science activities Parents' perception of school quality Parents' views on the importance of science learning Parents' reports on science career motivation Parents' general value of science Parents' personal value of science Parents' love of concern for environmental issues Parents' optimism regarding environmental issues

Because the contextual variables used in NAEP, TIMSS, and PISA are theoretically valid and reliable and supported by empirical evidence, Table 3.5 is an ideal pool of potential contextual variables for incorporation in a national educational assessment model.

3.4 Research Literature on Contextual Variables

The contextual variables used in NAEP, TIMSS, and PISA are known to have explained a certain percentage of variation in student outcomes in their own contexts (Lee, et al., 2007; Mullis, et al., 2004; OECD, 2007a). Guided by the contextual variables used in NAEP, TIMSS, and PISA, this section conceptualizes some select contextual variables and examines critically their relation with student achievement and contribution to quality education. The variables are examined as related to the four levels of the proposed national educational assessment model, namely, student-, classroom-, school-, and context-levels. Some variables feature at all four levels of the proposed model, and such variables are examined at student-level. Similarly, the variables common at student- and classroom-levels are examined at student-level.

The variables examined at classroom-, school-, and contextual-levels are known to have less direct influence on student achievement than the variables examined at the student-level. However, as assumed by the model, the variables at the higher levels facilitate the functioning of the variables at the lower levels, and thus the use of the variables at all levels of the proposed model for a national educational assessment programme should be of much interest and value to policymakers and other stakeholders of a nation's education system.

3.4.1 Educational Effectiveness Variables at Student-Level

Educational effectiveness variables at student-level comprise the following: gender; age; socio-economic status (SES); engagement; motivation; self-efficacy; self-regulated learning skills; homework; ICT; classroom management; and school climate. These variables are examined in terms of their association with student achievement and contribution to quality education.

3.4.1.1 Gender

Gender is an important variable in education (Clark, Lee, Goodman, & Yacco, 2008). Boys and girls differ in their academic achievements, though the differences are not always significant. Any significant achievement gap between boys and girls may require teachers, curriculum developers, policy-makers, and other stakeholders to design educational interventions to narrow the gap (Clark, Lee, et al., 2008; Weaver-Hightower, 2003).

Gender is widely used in explaining variations in student achievement, and most of the studies reported a consistent trend of girls outperforming boys academically (Clark, Lee, et al., 2008; Clark, Thompson, & Vialle, 2008; Lisle, Smith, & Jules, 2005; Younger & Warrington, 2007). However, researchers caution against the use of excessive gender specific (female oriented/appropriate or male oriented/appropriate) educational interventions to address the achievement gap (Weaver-Hightower, 2003; Younger & Warrington, 2006, 2007).

Researchers have also used gender to explain the effects of peer relationships (Crosnoe, Riegle-Crumb, Frank, Field, & Muller, 2008), co-education (Malacova, 2007; Younger & Warrington, 2006), academic cultures (Houtte, 2004a, 2004b), class attitudes (Gaer, Pustjens, Damme, & Munter, 2006), achievement orientations (Graham, Tisher, Ainley, & Kennedy, 2008), and subject selection (Cox, 2005) on student achievement. Close friends' achievement has superior influence over course mates' achievement on students' mathematics course-taking, indicating the potential use of peer relationships in identifying the risk factors associated with girls and boys dropping out of mathematics courses (Crosnoe, et al., 2008). Pupils with low prior attainment attending the schools with pupils of wide-ranging abilities are reported to benefit from attending single-sex classes co-education (Malacova, 2007; Younger & Warrington, 2006), indicating the possibility of using single sex-classes in co-education to improve student achievement. Boys' culture is reported as being less academically oriented than girls' culture, and the presence of girls in schools is found to improve pupils' academic culture in favour of boys (Houtte, 2004a, 2004b), indicating the role of gender in reducing gendered academic cultures. Class attitudes are known to affect the language achievement of boys greater than that of girls, and the effect is reported as strong among the same-sex class mates (Gaer, et al., 2006), signalling the prospect of using gender in fostering positive class attitudes. Traditional gender-stereotyped enrolment patterns are reported in mathematics and science subjects in Australia, with high-ability girls likely to choose the courses that are less difficult and low-ability boys likely to choose the courses that are more difficult (Cox, 2005). This finding highlights the potential use of gender to help enable student selection of ability-appropriate subjects.

Gender has been a regular focus of international assessment programmes. TIMSS 2007 noted that at Grade 4, on average, there was no gender difference in mathematics. However, at Grade 8, on average, girls had higher average achievement than boys in mathematics. PISA 2006 noted that, in general across the OECD countries, boys performed better than girls (OECD, 2007a). In addition, boys were more motivated, interested, and enjoyed the PISA mathematical literacy than girls. Further, the boys had less anxiety, higher self-efficacy, higher self-concept, and used more elaborate learning strategies (e.g., exploring how material relates to other contexts) than girls in learning the PISA mathematical literacy. However, girls used more control strategies (e.g., monitoring one's learning) in learning mathematics than boys.

In summary, gender has a wide range of implications in education. Data on gender are important for designing educational interventions to address various gender implications, like the ones discussed in this section. This suggested gender as a powerful factor, for a national educational assessment model, capable of assisting teachers, curriculum developers, policy-makers, and other stakeholders in designing educational programmes that engage all students.

3.4.1.2 Age

Children’s readiness for schooling and their ability to cope with cognitive, social, and emotional pressures in schools are widely researched in terms of their chronological age (Braymen & Piersel, 1987). In addition, some schools use children’s age as the basis for enrolment, grade retention, and grouping which have crucial implications on school resources, children’s opportunity costs in their adulthoods, and suitability of school curricula. However, researchers differ in their views on the roles of age in children’s academic achievement, social, and emotional development.

Students’ age has been studied in a range of contexts to explain student achievement: school readiness, academic redshirting (holding students out of scholastic competition with a view to enhancing the competitive advantage based on chronological maturity), grade retention, and school dropout. Researchers have reported a statistically significant, albeit small, achievement difference between the early entrants and the late entrants (Braymen & Piersel, 1987; Cameron & Wilson, 1990). However, in most cases, the difference is either found insignificant when students’ IQ is controlled (Kundert, May, & Brent, 1995) or is reported to diminish when students progressed to middle school (Hauck & Finch(Jr), 1993). Also, ‘academic redshirting’ is not associated with any advantage in achievement as a consequent of delaying school entry (Cameron & Wilson, 1990). Further, grade retention, which is commonly practised by schools to ameliorate a student’s achievement deficit, effectively makes the retained students older by the years retained than their fresh grade-mates. Jimerson (2001), in his meta-analysis of research literature published between 1990 and 1999 on the effects of grade retention on student outcomes, concluded that the grade retention has failed to improve student outcomes academically, socially, emotionally, and behaviourally. Grade retention is also reported as one of the most powerful predictors of student dropout prior to graduation (Guevremont, Ross, & Brownell, 2007; Jimerson, Anderson, & Whipple, 2002; Jimerson et al., 2006). Further, any short-term benefits associated with grade retention are known to dissipate across later grades (Jimerson, 2001; Jimerson, et al., 2002; Jimerson, et al., 2006). PISA 2006 also reported that student tracking and streaming at an early age is negatively associated with students’ SES and the proportion of repeaters at primary, lower secondary, and upper secondary schools (OECD, 2007a).

Researchers agree on the need to focus on instructional strategies, curriculum designs, socio-emotional problems of students, and remedial measures to address the needs of low achievers, instead of attributing the variations in student achievement to chronological age, grade retention, and redshirting (Braymen & Piersel, 1987; Cameron & Wilson, 1990; Hauck & Finch(Jr), 1993; Jimerson, 2001; Jimerson, et al., 2006; Kundert, et al., 1995).

In summary, students' chronological age has the potential to reveal important patterns in their social, emotional, and behavioural adjustments to school. Knowledge about such patterns will help teachers, curriculum developers, and policy-makers to design educational interventions to improve student achievement. Therefore, it is desirable for a national educational assessment model to have students' age as one of its variables.

3.4.1.3 Socio-Economic Status

A wide range of literature is available on the relation between SES and student achievement, with emphasis on the factors influencing the relation, and the utility of SES as a research tool in student achievement studies (Sirin, 2005; White, 1982; Wiggan, 2007). Sirin (2005) and White (1982) reported in their meta-analytic studies of the relation between SES and student achievement that a consensual definition of SES is hard to find. Nonetheless, Sirin (2005, p. 418) broadly defined SES as a description of "individual's or family's ranking on a hierarchy according to access to or control over some combination of some valued commodities such as wealth, power and social status". This definition of the SES is suitable for the use of SES in a national educational assessment model because it includes a range of attributes of individuals and families that make up social hierarchies.

The correlation between SES and student achievement is widely reported as ranging from moderate to strong (Caldas & Bankston(III), 1997; Lee, et al., 2007; Mullis, et al., 2004; OECD, 2004a, 2007a; Sirin, 2005; White, 1982). The wide-ranging correlation between SES and student achievement is attributed to the use of different units of analysis, measures of SES, types of academic measures, restriction of SES variable, type of SES components, and source of SES data by different researchers (Sirin, 2005; White, 1982).

Based on its relationship with student outcomes, SES is also used as a pointer for allocating educational resources to schools. When the achievement gap between certain groups of students is linked to their SES or other factors related to it, SES is reported to help policy-makers provide educational resources and support where most needed (Sirin, 2005). Further, the OECD (2007a) used the SES of students and schools as a criterion for assessing equity in, and accessibility to, educational resources. The OECD (2007a) hypothesized that students and schools who consistently perform well irrespective of their SES indicate equitable distribution of educational resources. Conversely, students and schools whose performance strongly depends on their SES are indicative of a large inequitable distribution of educational resources. Based on this hypothesis, the OECD (2007a) conclusively stated that it is possible to moderate the impact of the SES on student outcomes through equitable distribution of educational resources across schools. Schools can reduce the relation between SES and student achievement by providing equalizing experiences to students (Downey, Hippel, & Broh, 2004; OECD, 2007a; Sirin, 2005; White, 1982). Thrupp and Lupton (2006) noted that a focus on SES, as a potential means of addressing social injustice, should be put in the broader contexts of schools in order to provide relevant and strong foundational information for policy decision and practice.

In summary, diverse effects of SES on student outcomes, and the different ways in which researchers use SES to explain differential student outcomes clearly necessitate the use of SES as a powerful predictor variable of student achievement in a national educational assessment model.

3.4.1.4 Engagement

Engagement is widely viewed as a way of enhancing student achievement and motivation, of reducing dropout rates, and of enabling schools to develop students with the knowledge and skills required for solving complex problems in real-world situations and for pursuing life-long learning (Finn, 1993; Finn, Pannozzo, & Voelkl, 1995; Finn & Rock, 1997; Finn & Voelkl, 1993; Fredricks, Blumenfeld, & Paris, 2004; Hughes, Luo, Kwok, & Loyd, 2008). Engagement is conceptualized as a multi-dimensional construct, encompassing behaviour, emotion, and cognition, with all three overlapping one another (Finn & Voelkl, 1993; Fredricks, et al., 2004). Behavioural engagement is characterized as positive conduct towards school norms, participation in extracurricular activities, and active involvement in learning and academic tasks. Emotional engagement is characterized as affective ties to lessons, teachers, classmates, and school. Cognitive engagement is characterized as a clear investment in learning and self-regulation.

Engagement is affected by different factors associated with the school, classroom, and individual (Buhs, Ladd, & Herald, 2006; Fredricks, et al., 2004; Hughes, et al., 2008; Sullivan, McDonough, & Prain, n.d). For instance, at the school level, engagement is mediated by school size, school goals, availability of choice, participatory-approach to policy decisions and management. At the classroom-level, engagement is dependent on teacher support, positive peer relationships, classroom structure, autonomy support, and task characteristics. At the student-level, engagement is correlated to an individual’s need for relatedness and identity, autonomy, and competence. Greenwood (1991), in his longitudinal analysis of at-risk and non-risk student engagement, reported that low-SES students are less engaged with learning than high-SES students. This implies that SES is another important factor affecting student engagement.

The TIMSS 2007 assessed student engagement by using students' opinions about school safety, attitudes towards subjects (e.g., liking the school), educational expectations, and student absenteeism as variables (Mullis, et al., 2008). TIMSS 2007 reported that student absenteeism was a serious problem in many countries, and was negatively related with student achievement (Mullis, et al., 2008). In addition, students who liked their subjects, who felt safe in school, and who aspired to complete university education are reported to have performed better than students who did not like their subjects, who felt insecure in school, and who did not expect to complete university education.

The PISA 2003 and the PISA 2006 used students' self-efficacy, self-concept, subject anxiety, and the sense of belongingness to school as measures of student engagement (OECD, 2004a, 2007a), and reported that high self-efficacy, high self-concept, strong sense of belongingness, and low subject anxiety as correlates of high student achievement in mathematics and science.

Clearly, engagement is an important index of effective schooling and merits dedicated in-depth research of its own. However, some measures of engagement when incorporated in a national educational assessment model can provide indicators for interventions in an education system.

3.4.1.5 Motivation

Motivation is mostly studied in terms of individuals beliefs, values, goals, and actions (Eccles & Wigfield, 2002). Two popular motivation constructs are intrinsic or value motivation and extrinsic or instrumental motivation (Eccles & Wigfield, 2002; Tileston, 2010; Vansteenkiste, Timmermans, Lens, Soenens, & Broeck, 2008). The former refers to engaging in an activity for its own sake because it is enjoyable and gratifying, while the latter refers to engaging in an activity to obtain an outcome separable from the activity itself, such as a reward.

Researchers have related motivation to goal orientations. Intrinsic motivation is linked to mastery goals (learning or task orientation) and extrinsic motivation to performance goals (ability or ego orientation), claiming that mastery goals are positively correlated with enhanced conceptual learning and task persistence compared with performance goals (Schunk, 1991, 1996; Vansteenkiste, et al., 2008). A mastery goal is linked to adaptive response patterns as characterized by persistence in the face of failure, use of complex learning strategies, pursuit of difficult and challenging material and tasks, and belief in effort as a means of success (Ames, 1992; Ames & Archer, 1988; Bell & Kozlowski, 2002; Dweck, 1986; Hidi & Harackiewickz, 2000). Performance goals are linked to a maladaptive response pattern, characterized by greater propensity to withdraw from tasks in the face of failure, less interest in difficult tasks, a tendency to seek less challenging material and tasks where success is likely, and belief in ability as a prime cause of success or failure (Ames, 1992; Ames & Archer, 1988; Bell & Kozlowski, 2002; Dweck, 1986; Hidi & Harackiewickz, 2000). Furthermore, individuals with performance goals view cognitive ability as an entity which is a fixed, uncontrollable personal attribute, whereas individuals with mastery goals view an incremental theory about their ability, that is, ability is malleable and can be developed through effort and experience (Ames & Archer, 1988; Bell & Kozlowski, 2002; Dweck, 1986). Drawing on the reciprocal relation between motivation, goal orientations, and student achievement, researchers reported positive correlation between motivation, goal setting, and student achievement (Hattie, 2009; Tileston, 2010), indicating the potential use of motivation and goal orientations for improving student achievement.

In summary, students' ability, motivation, and goal orientations have been confirmed as powerful psychological constructs that influence teaching and learning in schools. Therefore, knowledge about how these constructs influence teachers and students has the potential to help stakeholders in developing educational interventions that foster appropriate experiences and perspectives of these constructs in teachers and students. One of the ways to acquire such knowledge is to incorporate these constructs into a national educational assessment model.

3.4.1.6 Self-Beliefs

Self-beliefs, consisting of self-concept and self-efficacy, have been keenly studied in the light of their associations with student achievement (Bandura, 1977; Bell & Kozlowski, 2002; Bong & Clark, 1999; Hattie, 2009; Pajares, 1996; Schunk, 1991; Valentine, DuBois, & Cooper, 2004).

Self-concept, as summarised by Bong and Clark (1999, p. 140), is "one's perception of the self that is continually reinforced by evaluative inferences and that it reflects both cognitive and affective responses". Drawing on this lead, students' self-concept is indicative of their affective responses to school education as the school provides students with evaluative norms and references for forming self-concept. Defining affect as a sense of positiveness or negativeness, Carver and Scheier (2005) convincingly noted that affects result in confidence or doubt, persistence or giving up, engagement or disengagement, and relief or anxiety. On the other hand, self-efficacy is defined as "people's judgment of their capabilities to organize and execute courses of action required to attain designated types of performances" (Bandura, 1986, p. 391).

In addition to their relationships with student achievement, self- concept and self-efficacy are known to associate with students' motivational and goal orientations (Bong & Clark, 1999; Hattie, 2009; Pajares, 1996; Schunk, 1991; Valentine, et al., 2004).

Bandura (1977) noted that self-efficacy affects a person's choice of activities, efforts, and persistence, resulting in mediating goal orientations and eventually motivation. Bong and Clark (1999) related self-concept (self-worth or self-esteem) to performance goal orientation and self-efficacy to mastery goal orientation. Bell and Kozlowski (2002) also reported that a mastery goal is positively related to self-efficacy. In addition, attributes of goals such as proximity, specificity, and difficulty are known to affect self-efficacy and motivation, with proximal goals promoting self-efficacy and motivation better than distal goals and with specific goals raising self-efficacy and motivation better than general goals (Bandura & Schunk, 1981; Schunk, 1991, 1996). Further, easier goals are known to enhance self-efficacy at the initial stages of skill acquisition; while difficult goals are known to enhance self-efficacy in the course of skill development (Bandura & Schunk, 1981; Schunk, 1991, 1996). These findings support the views that self-efficacy is acquired from sources such as performance accomplishments, vicarious experience, verbal persuasion, and emotional arousal (Bandura, 1977). In addition, students' goal orientations are mediated by their perception of classroom structures such as the ways in which learning tasks are developed, the type of authority exercised, and the kinds of evaluation and recognition exercised and followed (Ames, 1992; Ames & Archer, 1988). Pajares (1996) authoritatively asserted that these sources when properly moderated will assist in developing students' self-efficacy, suggesting self-efficacy as a malleable trait.

Similar to the role of students' self-efficacy in their learning outcomes, teachers' satisfaction with their job is also reported to be positively correlated with their self-efficacy, indicating the need to study teachers' self-efficacy as much as the need to know about students' self-efficacy (Caprara, Barbaranelli, Borgogni, & Steca, 2003; Goddard & Goddard, 2001; Goddard, Hoy, & Hoy, 2000; J. A. Ross, 1992; Tschannen-Moran, Hoy, & Hoay, 1998).

In summary, knowledge about self-beliefs of students and teachers has the potential to guide stakeholders in developing educational interventions for promoting motivation and facilitating goal-orientations to improve learning and teaching in schools, making self-beliefs a desirable variable in a national educational assessment model.

3.4.1.7 Anxiety

Anxiety is generally associated with a condition of emotion that is characterised by fear, uncertainty, helplessness, and dread (Ashcraft & Moore, 2009; Geist, 2010; Hembree, 1990; Ma, 1999). In addition to its relationship with student achievement, anxiety is known to elicit a range of negative affects in students (Carver & Scheier, 2005).

Anxiety and student achievement are known to relate inversely with each other, with high anxiety relating to low student achievement (Ashcraft, 2002; Ashcraft & Moore, 2009; Hembree, 1990; Ma, 1999; OECD, 2004a). Anxiety toward a school subject (e.g., mathematics) is also known to relate with students' negative attitude toward the subject and their tendencies to avoid the subject (Ashcraft & Moore, 2009; Hembree, 1990).

Given such negative consequences of students’ subject anxiety on their learning outcomes, a range of solutions to prevent or reduce subject anxiety (e.g., mathematics) has been presented by researchers, emphasising the tractable nature of anxiety (Furner & Berman, 2003; Geist, 2010; Hellum-Alexander, 2010; Hembree, 1990; Ho et al., 2000; Newstead, 1998; Norwood, 1994).

In summary, knowledge about students’ anxiety toward their subject (e.g., mathematics) and the tractable nature of anxiety has the potential to guide stakeholders in developing educational interventions for preventing or reducing anxiety, making anxiety a desirable variable in a national educational assessment model.

3.4.1.8 Students’ Preferences for Learning Environments

Students’ preferences for learning environments are broadly categorized into competitive and cooperative learning environments, and they are reported to have wide-ranging implications for students’ development.

Stapel and Koomen (2005) noted that students can either adopt a competitive or a cooperative learning preference based on how they view themselves with respect to who they compare with. Stapel and Koomen (2005) persistently claimed that these views form the basis of students’ self-evaluation. Self-evaluation from either competitive or cooperative learning preferences has consequences on students’ motivational and goal orientations, with the competitive learning preferences related to instrumental motivation and performance goals, and the cooperative learning preferences related to intrinsic motivation and mastery goals (Covington & Omelich, 1984). Further, students’ preferences for learning environments have been linked to teaching strategies, with competitive learning preference linked to structured teaching strategies and cooperative learning preference linked to constructivist teaching strategies (Covington & Omelich, 1984; Ediger, 1996).

In summary, knowledge about students’ preferences for learning environments may help Bhutanese educators to develop educational interventions to improve student outcomes, making students’ preferences for learning environments a useful factor for a national educational assessment model.

3.4.1.9 Self-Regulated Learning Skills

One of the major goals of schooling is the promotion of self-regulatory learning skills in students so that they become capable of monitoring, controlling, and evaluating their own learning. Puustinen and Pulkkinen (2004), in their review of models of self-regulated learning, highlight the importance of self-regulatory skills by referring to them as necessary preconditions for meaningful, effective, and successful life-long learning.

Researchers have reported a substantial correlation between student performance and the use of self-regulated learning strategies (Eccles & Wigfield, 2002; OECD, 2004a, 2007a; Pintrich & De Groot, 1990; Purdie & Hattie, 1996; Schunk, 1996; Zimmerman, 1990; Zimmerman, Bandura, & Martinez-Pons, 1992). Self-regulated learning strategies have also been related to self-efficacy, goal orientations, and test anxiety (Eccles & Wigfield, 2002; Pintrich & De Groot, 1990; Wigfield & Eccles, 2000; Zimmerman, 1989, 1990; Zimmerman, et al., 1992). Mastery goal orientation and self-efficacy are strongly correlated to self-regulated learning (Pintrich & De Groot, 1990). Self-regulation, self-efficacy, and test anxiety have been found to be the best predictors of student outcomes (Pintrich & De Groot, 1990).

A self-regulated student is characterized as one who is metacognitively, motivationally, and behaviourally active in his/her own learning process (Pintrich & De Groot, 1990; Zimmerman, 1989, 1990; Zimmerman & Martinez-Pons, 1988). Zimmerman (1990) compiled the attributes of metacognitive, motivational, and behavioural components of self-regulated learning. In terms of metacognitive processes, self-regulated learners plan, set goals, organize, self-monitor, and self-evaluate. In terms of motivational processes, self-regulated learners report high self-efficacy, self-attributions, and intrinsic interest. In terms of behavioural processes, self-regulated learners select, structure, and create environments that optimize learning. Further, Zimmerman (1989, 1990) and Zimmerman and Martinez-Pons (1988) and Pelgrum (2001) identified 14 self-regulated learning strategies as follows: self-evaluation, organization and transformation, goal-setting and planning, information seeking, record keeping and self-monitoring, environmental structuring, giving self-consequences, rehearsing and memorizing, seeking social assistance, and reviewing records. They convincingly claim that these strategies represent a range of behaviours that students engage in to regulate their personal functioning, academic behavioural performance, and learning environment.

Self-regulated learning strategies are also known to depend on learning environments. Purdie and Hattie (1996) studied patterns in the use of self-regulated learning strategies by students in terms of their cultural backgrounds and noted that the usage varied across cultures—Japanese students used memory strategies significantly more than Australian students. Similarly, Boekaerts and Niemivirta (2005) authoritatively noted that traditional school settings—where teachers set learning goals for students, provide students with declarative and procedural knowledge, and make students responsible for finding means to comprehend, store, and activate that knowledge—actually reduce the opportunities for students to self-regulate their own learning. All this indicates that self-regulated learning strategies can be fostered by creating enabling learning environments that invite students to organize and regulate their own learning.

In summary, self-regulated learning is an important skill that students need to develop not only because it is correlated with student outcomes, but also because it is a fundamental building unit of knowledge and skills characteristic of the twenty-first century knowledge economy; therefore, it merits a place in a national educational assessment model as one of its variables.

3.4.1.10 Homework

The relation between homework and student achievement has been of a great interest to educational researchers (Brock, Lapp, Flood, Fisher, & Han, 2007; Cooper, Robinson, & Patall, 2006; Epstein & VanVoorhis, 2001; Muhlenbruck, Cooper, & Nye, 2000). Homework is widely understood as learning tasks assigned to students by school teachers that are carried out outside school hours (Cooper, 1989; Cooper, et al., 2006). According to this concept of homework, in-school guided study, home study courses, and extra-curricular activities are not considered as homework (Cooper, et al., 2006).

Homework is widely reported as a correlate of student achievement (Brock, et al., 2007; Cooper, 1989; Cooper, Lindsay, & Nye, 1998; Cooper, et al., 2006; Muhlenbruck, et al., 2000; Mullis, et al., 2008; OECD, 2007a). Cooper et al. (2006), in their meta-analysis of studies on the relation between homework and student achievement published between 1987 and 2003, reported that homework is consistently related to student achievement. In addition, the correlation between homework and student achievement is reported as strong at secondary school level, moderate at lower secondary school level, and weak at elementary school level. The optimal time for homework is noted to be in the range of one to 10 hours per week for secondary school and one to 5 hours per week for elementary school (Cooper & Valentine, 2001).

Homework is also known to have both positive and negative effects on student outcomes. The positive effects of homework comprised immediate student achievement and learning, long-term academic benefits, non-academic benefits (e.g., self-discipline), and parental and family benefits (Cooper, et al., 2006). Drawing on a range of studies on interaction between homework and parents' involvement, Hoover-Dempsey (2001) noted that homework engages parents in their children's studies as follows: providing study rooms, role modelling study habits, offering instructions and reinforcements, and facilitating self-regulated learning skills. Homework is also known to facilitate peer-learning (Epstein & VanVoorhis, 2001). On the other hand, satiation effects, parental interference, cheating, denial of leisure time and access to community activities, and widening the achievement gap between high and low achievers are reported as some negative effects of homework (Cooper, et al., 2006). Homework is also reported to accentuate social inequities, with students from low SES families unfairly challenged by their homework because of dissimilarities in their home and instructional languages or because of non-availability of educational resources at their home (Brock, et al., 2007; Cooper, 1989).

In summary, because homework has a range of positive and negative effects on student outcomes, using it as a variable in a national educational assessment model has the potential to provide useful findings to students, teachers, and schools.

3.4.1.11 Information and Communications Technology (ICT)

Lim (2002, p. 412) conceptualized ICT as “a mediational tool, incorporated within learning environments with authentic goals and purposes for students, and settings that are explicitly interpreted with other experiences of knowing and understanding as they get organised at other times”. Ruthven, Hennessy, and Brindly (2004) aptly claimed that ICT as a mediational tool can expand beyond ICT- specialist courses and projects to everyday practice of mainstream schooling.

Researchers have reported a range of positive influences of ICT on student outcomes. ICT is known to add expediency and productivity to students’ work, foster students’ independence from teachers, encourage peer learning among students, broaden students’ information resource base, and maintain the currency of students’ learning activities (Ruthven, et al., 2004). ICT is also reported to improve students’ self-regulated learning strategies (Ilomaki & Rantanen, 2007; Lim, 2002). These findings are corroborated by the results from TIMSS and PISA that reported positive correlation between students’ use of ICT and their academic achievement level (Mullis, et al., 2008; OECD, 2004a).

Besides its association with student outcomes, ICT is also associated with teaching pedagogies. Drent & Meelison (2008) reported that the use of ICT by teacher educators is positively influenced by student-oriented pedagogical approaches, positive ICT attitudes, computer experience, and personal entrepreneurship; indicating the potential of ICT to drive student-oriented teaching.

Investment in ICT is also reported to be on the rise because of its diverse roles. Automating routine jobs, facilitating timely and competent participation in labour markets, enhancing opportunities for life-long learning, improving teaching and learning, and developing transparency, accountability, and efficiency are some roles of ICT that attract countries to invest in it (OECD, 2005b). The underlying reason for investment in ICT is its potential to be an indispensible tool for successfully coping up with the twenty-first century world of information and knowledge economies. However, Pelgrum (2001) noted that much has to be learnt about the ways in which ICT is pursued by schools and systems.

Given the importance of ICT in education as above, knowledge about how ICT is used by students, teachers, and principals will shape ICT-related policy decisions. One way of obtaining such knowledge is by including ICT as a variable in a national educational assessment model such as the one that this study seeks to develop for Bhutan.

3.4.1.12 Classroom Management

Classroom management is a common topic that features in pre-service teacher preparation courses and in in-service professional development programmes. Classroom management is widely studied in terms of teacher and student behaviours and physical arrangement of the classroom (Doyle, 1980; Marzano, 2003a, 2003b; Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008).

Effective classroom management is known to increase student outcomes. Cothran, Kulinna, and Garrahy (2003) claimed that an effective classroom is indispensible for a safe and functional class, and for all other teaching and learning activities in a class to be successful, indicating the role of classroom management in student achievement. Marzano (2003a) also reported effective classroom management as a correlate of greater student engagement and higher student achievement.

Researchers have identified various features of classroom management. In his meta-analysis of studies on classroom management, Marzano (2003a, p. 8) identified four general elements of effective classroom management: “rules and procedures, disciplinary interventions, teacher-student relationships, and ‘mental set’, noting that an effective classroom management is most likely when all four components function in concert. Simonsen et al. (2008, p. 352) identified five categories of empirically-supported critical features of effective classroom management: “physical arrangement of classroom; structure of classroom environment; instructional management; procedures designed to increase appropriate behaviours; and procedures designed to decrease inappropriate behaviours”. They also presented practices that indicate the existence of their five critical features of classroom management as shown in Table 3.6.

Table 3. 6. Effective Classroom Management Practices

Five Empirically-supported Critical Features of Effective Classroom Management	Evidence-based Practices
Maximise Structure and Predictability	High classroom structure (e.g., amount of teacher directed activity) Physical arrangement that minimises distractions (e.g., walls, visual dividers) and crowding
Post, Teach, Monitor, and Reinforce Expectations	Post, teach, review, and provide feedback on expectations Active supervision
Actively engage students in	Rate of opportunities to respond

observable ways	Response cards Direct instruction Computer assisted instruction Class peer tutoring Guided notes
Use a Continuum of Strategies to Acknowledge Appropriate Behaviours	Specific/contingent praise Class-wide group contingencies Behavioural contracting Token economies
Use a Continuum of Strategies to Respond to Inappropriate Behaviours	Error corrections Performance feedback Differential reinforcement Planned ignoring plus contingent praise and/ or instruction of classroom rules Response cost Time out from reinforcement

Simonsen et al. (2008) authoritatively stated that teachers who are knowledgeable of, and report practising 80% of, the classroom management practices depicted in Table 3.6 are more likely to achieve effective classroom management than their colleagues who report practising less than 60% of the same.

Like the above features of effective classroom management, researchers have also identified ways of achieving effective classroom management from the perspectives of students and teachers. Students are known to perceive classroom management as effective when they experience announced and clearly consistent standards for expectations and consequences, and caring and respectful relationships with their teachers (Cothran, et al., 2003). Similar to the students' perception of effective classroom management, teachers equate effective classroom management with consistent classroom rules, providing reinforcements for appropriate behaviours, responding to class disruption, and dealing with chronic classroom offenders (Akin-Little, Little, & Laniti, 2007; Emmer, Evertson, Sandford, Clements, & Worsham, 1984; Little & Akin-Little, 2008). Drawing on Stage and Quiroz's

(1997) meta-analysis of the literature on interventions designed to decrease disruptive classroom behaviours, Marzano (2003a, 2003b) identified four disciplinary techniques: reinforcement, punishment, combination of reinforcement and punishment, and no immediate consequence. The effect sizes of these four disciplinary techniques reported by Stage and Quiroz (1997) are shown in Table 3.7.

Table 3. 7. Four Disciplinary Techniques to Improve Classroom Management

Intervention	Average Effect Size	Number of Effect sizes	SD
Reinforcement	-0.86	101	0.58
Punishment	-0.78	40	0.47
Reinforcement and punishment	-0.97	12	0.89
No immediate consequence	-0.64	70	0.54

An immediate inference from Table 3.7 is that while the disciplinary technique involving both reinforcement and punishment is most effective, the technique of no immediate consequence is least effective. This implies that students need to be appraised of their appropriate and inappropriate behaviours by providing relevant reinforcements and punishments.

In summary, effective classroom management is indispensible for effective and meaningful teaching and learning to occur in classroom. Therefore, it is essential that policy-makers and other stakeholders in a nation’s education system are informed on their teachers’ knowledge and skills of effective classroom management, and use such information to develop effective educational interventions. By including classroom management as a variable in a national educational assessment model, teachers’ knowledge and skills of effective classroom management can be evaluated, and the resulting information can be used to guide policy decisions and educational interventions.

3.4.1.13 School Climate

School climate is broadly defined as the collective beliefs, values, and attitudes that influence interactions between students, teachers, and administrator, resulting in a set of parameters of acceptable behaviours and norms for schools (Hoy & Hannum, 1997; Koth, Bradshaw, & Leaf, 2008). In other words, school climate is a multi-dimensional concept, with the assumption that an enabling school climate of an orderly respectful atmosphere could lead to higher student achievement (Brand, Felner, Shim, Seitsinger, & Dumas, 2003; Marzano, 2001; Scheerens, 1990, 1997, 2000; Scheerens, Glass, & Thomas, 2003).

Meta-analytic studies of school effectiveness have shown that school climate is positively correlated with student achievement (C. S. Anderson, 1982; Marzano, 2001; Sammons, et al., 1995; Scheerens, 2000). Other independent studies also reported evidences of correlation between school climate and student achievement (Brand, et al., 2003; Hoy & Hannum, 1997; Macneil, Prater, & Bush, 2007; Mullis, et al., 2008; OECD, 2004a). School climate is also associated with other student outcomes. School climate is linked to: student engagement (Finn & Voelkl, 1993); student misconduct, aggression, and behavioural problems (Wilson, 2004); and drug abuse and delinquent behaviour (Battistich & Hom, 1997; Battistich, Solomon, Kim, Watson, & Schaps, 1995). These findings underscore a range of influences of school climate on students and their growth trajectories, including academic achievement.

School climate is malleable, and knowing the factors that change it will help schools to make it more conducive to teaching and learning. Structured rules and regulations, punishment and reward system, absenteeism and drop-out, good conduct of pupils, and achievement orientation are reported as predictors of school climate (Marzano, 2001; Scheerens, et al., 2003). In addition, Koth et al. (2008) reported school size, teacher turnover, teacher behaviours, class size, concentration of students with behavioural problems, and students' race and gender as correlates of school climate.

Given that school climate has a wide range of influences on student outcomes, it is desirable to incorporate school climate as a variable in a national educational assessment model. Data on school climate can provide a valuable insight into the overall orderliness or 'temperature' of schools so that appropriate interventions can be developed if necessary.

3.4.2 Educational Effectiveness Variables at Classroom-Level

The following educational effectiveness variables at classroom-level are examined in this section: professional development; appraisal and feedback; and teaching effectiveness components. Other classroom-related variables like self-efficacy, classroom management, school climate, homework, and ICT have already been examined in earlier sections, or in the case of teacher demographics will be examined in a later section as part of school resources

3.4.2.1 Professional Development

The OECD (2009a, p. 49) defines professional development as "activities that develop an individual's skills, knowledge, expertise, and other characteristics as a teacher". The activities include course/workshops, education conferences or seminars, qualification programmes, observation visits to other schools, participation in a network of teachers, individual or collaborative research, and mentoring or peer observation and coaching. These activities can occur in a formal or an informal setting.

Professional development helps teachers enhance their content knowledge and develop their teaching strategies and skills, thereby making it an essential component of systemic reform to develop teachers' capacity to teach to high standards. Researchers generally agree that professional development focussed on specific instructional practices increases teachers use of those practices in the classroom (Desimone, Garet, Yoon, & Birman, 2002; Yoon, Garet, Birman, & Jacobson, 2006). Starkey et al. (2009) noted that professional development also supports major educational changes and reforms that have an impact on teaching practice, indicating the significance of professional development in educational interventions.

Researchers increasingly agree on emerging characteristics of high quality professional development. A high quality professional development: is content-oriented and focussed on how students learn content; provides teachers with in-depth, active learning opportunities; provides teachers with opportunities to engage in leadership roles; has the collective participation of groups of teachers from the same school, grade, or departments; and is relevant to teachers' goals and teaching experiences (Desimone, et al., 2002; Starkey, et al., 2009; Wayne, Yoon, Zhu, Cronon, & Garet, 2008; Yoon, et al., 2006).

Given that professional development affects teachers' instructional practices and supports major educational reforms and that certain characteristics have to be met for professional development to be effective in its purposes, the incorporation of professional development in a national educational assessment model will provide valuable insight into the patterns of teachers' participation in such professional development. This insightful knowledge can assist policy makers in planning professional development programmes for Bhutanese teachers.

3.4.2.2 Appraisal and Feedback

The OECD (2009a, p. 141) states that "teacher appraisal and feedback occurs when a teacher's work is reviewed by either the school principal, an external inspector or the teacher's colleagues". Teacher appraisal and feedback are commonly viewed as a mechanism for encouraging professional learning and growth in teachers, identifying opportunities for additional support for teachers, and providing a measure of accountability with the view to fostering sustained teacher development (Ministry of Education, 2010).

The OECD (2009a) reported that teacher appraisal and feedback have a strong positive influence on teachers and their work, with the influences resulting in job satisfaction, job security, and high standard teaching practices among teachers.

Considering that teacher appraisal and feedback are a part of the education system in many countries, including Bhutan, knowledge about how teacher appraisal and feedback are conducted in schools is essential for policy makers. Smith (1995) reported that the effectiveness of a teacher appraisal and feedback system depends on organizational commitment, adequate resources, identifying and communicating appraisal areas, consistency of good practice, opportunities for adequate follow up, and positive views on the outcomes of the appraisal. Similarly, other researchers suggest that teacher appraisal should focus on teacher development rather than focussing on teacher accountability and that the purposes of teacher appraisal should be clearly identified (Casey, Gentile, & Biger, 1997; Gratton, 2004).

A national educational assessment programme can provide information about the state of teacher appraisal and feedback in schools. A national educational assessment programme can also collect teachers' perspectives of their appraisal and feedback. Such information has the potential to help policy makers in improving teacher appraisal and feedback systems to enhance their impact on better teaching and learning in schools, making teacher appraisal and feedback a desirable variable for a national educational assessment model.

3.4.2.3 Teaching Effectiveness Components

Researchers in teaching effectiveness have sought to identify the most effective teaching components, but concluded that there is no particular instructional strategy that works best across all situations and all domain-specific knowledge (Harris, 1998; Marzano, et al., 2000). Consequently, researchers have identified generic instructional strategies that are associated with effective teaching and learning when used together, depending on the classroom situation and domain specific knowledge (Harris, 1998).

Teachers' knowledge of different instructional variables and their ability to use them effectively during lesson sessions is widely purported to enhance student learning (Marzano, et al., 2000; Sammons, et al., 1995; Seidel & Shavelson, 2007). In addition, Marzano (2003b) convincingly notes that expert teachers learn a range of instructional strategies together with the knowledge of using them in the manner that drives effective student learning.

Fraser et al. (1987, p. 157) identified 26 instructional variables associated with effective teaching and learning in their classic synthesis of educational productivity research studies published between 1970 and 1987. These instructional variables are listed in Table 3.8.

Table 3. 8. The Fraser List of Teaching Variables with Effect Sizes

Instructional Variable	Effect Size	Instructional Variable	Effect Size
Reinforcement	1.17	Individualized mathematics	0.32
Acceleration	1.00	New science curricula	0.32
Reading training	0.97	Teacher expectations	0.31
Cues and feedback	0.97	Computer assisted instruction	0.28
Science mastery	0.81	Sequenced lessons	0.24
Cooperative programs	0.76	Advanced organizers	0.24
Reading experiments	0.60	New mathematics curricula	0.23
Adaptive instruction	0.57	Inquiry biology	0.18
Tutoring	0.45	Homogeneous groups	0.16
Individualized science	0.40	Programmed instructions	0.10
Higher-order questions	0.35	Class size	-0.03
Diagnostic prescription	0.34	Mainstreaming	-0.09
Individualized instruction	0.33	Instructional time	-0.12

Fraser et al. (1987) plausibly inferred from the effect sizes shown in Table 3.8 that effective teaching and learning is the combined result of several factors, not as a result of any single factor. For instance, an effect size of approximately 3.8 is reported when all the instructional variables, as listed in Table 3.8, are used simultaneously (Fraser, et al., 1987, p. 160). Mainstreaming and class size have small effect sizes and are negative, implying that their influence on student outcomes will result in a small negative return. These will also normally require more teachers and more classrooms, and may draw more money and effort from the factors with large effects (Fraser, et al., 1987).

Scheerens and Bosker (1997, p. 305), in their research synthesis of school effectiveness studies, identified eight instructional variables associated with effective teaching and learning. These eight variables are listed in Table 3.9.

Table 3. 9. The Scheerens List of Eight Instructional Variables with Effect Sizes

Instructional Variables	Effect Size
Reinforcement	0.58
Feedback	0.48
Cooperative learning	0.27
Differentiation/adaptive instruction	0.22
Time on task/homework	0.19
Structured teaching	0.11
Opportunity to learn	0.09
Homework	0.06

In Table 3.9, reinforcement has the highest effect size and homework has the lowest effect size. A plausible explanation for the observed differences in effect sizes is that the instructional variables that are closer to the operating core of learning and instruction have greater impact on student outcomes than the distal instructional variables that do not influence the immediate interactive experience of students (Scheerens & Bosker, 1997).

Marzano (2001, p. 63), in his synthesis of the studies on teaching effectiveness from the past 40 years, identified nine instructional variables that he considered as the most effective teaching variables in optimizing student learning. Also, the same instructional variables are revisited by Marzano (2003b, p. 80) in his more recent synthesis of the studies on school effectiveness from the past 35 years. These instructional variables are listed in Table 3.10.

Table 3. 10. The Marzano List of Nine Instructional Variables with Effect Sizes

Instructional Variable	Effect Size	Standard Deviation
Identifying similarities and differences	1.61	0.30
Summarizing and note taking	1.00	0.50
Reinforcing effort and providing recognition	0.80	0.35
Homework and practice	0.77	0.36
Non-linguistic representations	0.75	0.40
Cooperative learning	0.73	0.40

Setting goals and providing feedback	0.61	0.28
Generating and testing hypotheses	0.61	0.79
Activating prior knowledge	0.59	0.26

In Table 3.10, the instructional variables differ in their effect sizes, highlighting the absence of a single instructional strategy that is capable of working equally well in all situations. Further, Marzano et al. (2000) noted that the effectiveness of instructional strategies are mediated by prior student knowledge, teachers' knowledge and skill to apply the instructional strategies effectively, and contextual factors such as class size and grade level.

Seidel and Shavelson (2007, p. 473), in their meta-analytic study of teaching effectiveness studies of the past 10 years, identified seven instructional variables which are associated with effective teaching and learning. These variables are listed in Table 3.11.

Table 3. 11. The Seidel List of Seven Teaching Variables with Effect Sizes

Instructional Variable	All Outcomes	Learning Processes	Motivational-Affective	Cognitive
	Effect Size	Effect Size	Effect Size	Effect Size
Time for learning	0.04	0.14	0.12	0.03
Organization of learning	0.01	0.01	0.06	0.00
Social context	0.04	-0.03	0.01	0.05
Goal setting and orientation	0.03	0.09	0.07	0.02
Execution of learning activities				
Social/direct experiences	0.01	0.11	0.13	0.00
Basic processing	0.02	0.05	0.08	0.01
Domain-specific processing	0.21	0.16	0.21	0.22
Evaluation of learning	0.01		0.00	0.02
Regulation/monitoring	0.02	0.05	0.08	0.01

A plausible inference from Table 3.11 is that while domain specific activities have the highest teaching effect on student outcomes, the teaching effects differ across different student outcomes. This underlines the inclination of researchers in teaching effectiveness to concentrate more on global aspects of teaching and studying teaching patterns in preference to focussing on a particular teaching strategy (Seidel & Shavelson, 2007). In addition, salient in Table 3.11 is the consistently small effects of individual teaching variables. A plausible explanation offered by Seidel and Shavelson (2007) for this observation is the aggregation of effect sizes over different teaching variables and student outcomes. For instance, the effects on cognitive outcomes are consistently lower than the effects on the other two outcomes, indicating that the "overall low effect sizes are due to low effects of teaching on cognitive outcomes" (Seidel & Shavelson, 2007, p. 472).

Hattie (2009), in his study of 31 meta-analyses, reported an average effect of d= 0.49, SE=0.049 of teacher-related factors on student achievement, indicating that a one standard deviation increase in teacher effectiveness should increase student achievement gains by about one-half of a standard deviation. Table 3.12 shows the various teacher-related factors and their corresponding effect sizes as reported by Hattie (2009, p. 109).

Table 3. 12. The Hattie List of 10 Teacher-Related Factors with Effect Sizes

Teacher-Related Factors	Effect Size	Standard Deviation
Teacher effects	0.32	0.020
Teacher training	0.11	0.044
Microteaching	0.88	-
Teacher subject matter knowledge	0.09	0.016
Quality of teaching	0.44	0.060
Teacher-student relationships	0.72	0.011
Professional development	0.62	0.034
Expectations	0.43	0.081
No labelling students	0.61	
Teacher clarity	0.75	
Total	0.49	0.049

As shown in the Table 3.12, except for teacher effects and teacher training, all teacher-related factors have high effect sizes, indicating their potential to influence student achievement. The effect size of teacher effects is medium and that of teacher training is small. Hattie (2009) attributed the medium effect size of teacher effects to a difference in school SES with teacher effects being much larger in low SES schools than in high SES schools. Likewise, the lack of common standards for teacher education, little use of research-based knowledge in teacher education, and high propensity of student teachers to revert to the teaching approaches of their favourite teachers from their own school days are emphasized as some reasons for low effect size of teacher training.

In summary, instructional components have effect sizes ranging from mild to strong. By incorporating teaching variables in a national educational assessment model, data can be collected on these variables for informing policy decisions and for developing interventions to maximise student outcomes.

3.4.3 Educational Effectiveness Variables at School-Level

The following educational effectiveness variables at school-level are examined in this section: school policies and practices related to school admittance, within school ability grouping; school autonomy; parental involvement; educational leadership; and school resources. School climate, as an educational effectiveness factor at school-level has been examined in the section on educational effectiveness factors at student-level. Similar to educational effectiveness factors at student-and classroom-levels, educational effectiveness factors at school-level are examined in terms of their association with student achievement and contribution to overall quality education.

3.4.3.1 School Policies and Practices related to School Admittance

School admittance policies are commonly grouped into residence-oriented, grade-oriented, and a hybrid of the two (OECD, 2007a; Soderstrom & Uusitalo, 2010). Residence-oriented school admittance policies consider students' residential proximity to school as the main criterion for admitting students into the school, while grade-oriented school admittance policies consider students' grades as the main criterion for the same purpose.

These school admittance policies and practices influence student achievement, school choice, student mobility, and social segregation (OECD, 2007a; Soderstrom & Uusitalo, 2010). Residence-oriented admittance policies segregate students in terms of their SES and neighbourhood, widening the gap between high- and low-performing schools because of limited student mobility from one neighbourhood school to another (Soderstrom & Uusitalo, 2010). On the other hand, grade-oriented school admittance policies facilitate student mobility from one school to another school as determined by their academic grade, enabling high-performing students to move from low-performing schools to high-performing schools. Grade-oriented school admittance policies also facilitate school choice for both parents and students. However, grade-oriented school admittance policies can result in segregation of fast learners from slow learners with the former attending high-performing schools and the latter attending low-performing schools (Soderstrom & Uusitalo, 2010). Similarly, grade-based school admittance policies are also known to widen the gap between families along their SES profiles, with the families of higher SES profiles living in or moving to places where high-performing schools are located. Open school admittance policies are known to enhance student achievement, student and teacher relationships, and student satisfaction at school (Cullen, Jacob, & Levitt, 2005; Gibbons, Machin, & Silva, 2008; Lavy, 2010). This is particularly the case for students in high-performing schools.

In summary, school admittance policies are associated with student achievement, school choice, student mobility, and social segregation. Therefore, it has high stakes implications for parents and students, which makes it desirable to consider as an educational effectiveness variable in a national educational assessment model.

3.4.3.2 School Policies and Practices related to Within School Ability Grouping

Ability grouping is a way of assigning students to a homogeneous group, with the group homogeneity determined by students' academic performance as measured by their examination scores or classroom performance. Ability grouping has been a subject of much debate among educationists, with some arguing that grouping students into a homogeneous group enables teachers to tailor their instructions to the group, some arguing that a heterogeneous grouping facilitates peer learning among students of different abilities, and some arguing that homogeneous ability grouping causes educational inequity among various student subgroups (Betts & Shkolnik, 2000; Gamoran, 1992; Hanushek & Wobmann, 2006; Hattie, 2009; Slavin, 1990).

Similar to the debate on its role in facilitating tailored classroom instruction and disabling peer learning, the relation of ability grouping with student achievement is inconclusive. Betts and Shkolnik (2000) found little difference in the average mathematics achievement growth of students in the schools that practised ability grouping and the schools that did not, and they attributed the absence of difference to the usage of school resources. Betts and Shkolnik (2000) observed similar allocation of school resources (teacher education, teacher experience, and class size) between schools that claimed to practise ability grouping and those that claimed not to practise ability grouping. Besides its lack of a consistent positive relationship with student achievement, ability grouping at school level is reported to exacerbate educational inequality between various student subgroups (Gamoran, 1992; Hanushek & Wobmann, 2006; Hattie, 2009).

Considering the inconclusive nature of the relation between ability grouping and student achievement, it will be of interest to stakeholders in the Bhutanese education system to know how schools practise ability grouping. Such knowledge has the potential to assist the stakeholders in developing educational interventions that maximise the positive effect of ability grouping, making ability grouping a desirable educational effectiveness variable in a national educational assessment model.

3.4.3.3 School Policies and Practices related to Autonomy

School autonomy entails the responsibilities devolved to schools by central educational authorities to improve the quality of education. Researchers trace the emergence of school autonomy in the ideals of democratic participation by stakeholders in school activities and efficient management of school funds, human resources, and school curriculum (Caldwell, 2005; Eurydice, 2007; Grauwe, 2005; Maslowski, Scheerens, & Luyten, 2007; West, Allmendinger, Nikolai, & Barham, 2010).

The relation between school autonomy and student achievement is reported to be inconclusive, though schools with more autonomy are known to perform better than schools with less autonomy (Fuchs & Wobmann, 2007; Maslowski, et al., 2007; OECD, 2010; West, et al., 2010; Wobmann, 2007). Maslowski et al. (2007) reported that schools' autonomy in personal management resulted in improved reading literacy, but stated that the improvement disappeared when student composition was considered. Grauwe (2005), acknowledging the inconclusive relation between school autonomy and student achievement, suggested that school autonomy should be supported by strategies to develop capacities of schools, principals, teachers, and communities with clear foci on quality improvement and concern for equity for it to be effective. School autonomy is also more positively related to student achievement in education systems that have national examinations (Fuchs & Wobmann, 2007; Wobmann, 2007).

In summary, considering the confusion surrounding the relation between school autonomy and student achievement, data on how school principals and teachers view various levels of autonomy in their responsibilities have potential to guide policy decisions related to school autonomy and so would be a desirable variable in the national educational assessment model.

3.4.3.4 School Policies and Practices related to Parental Involvement

Parental involvement in their children's education is increasingly viewed as a family educational resource input to improve student outcomes (Fan & Chen, 2001; Hoover-Dempsey & Sandler, 1997; Jeynes, 2003; Yan & Lin, 2005). Grolnick and Slowiaczek (1994) described three types of parental involvement in children's education: behaviour (e.g., participation in activities in school and at home); cognitive-intellectual (e.g., providing intellectually stimulating resources); and personal (e.g., knowing about their child's activities and preferences and general knowledge of this area). Epstein (1995) described six types of parental involvement: parenting practices at home; school-home communication; volunteering to undertake school activities, involvement in learning at home; involvement in school decision-making; and collaborating with the wider community. Hoover-Dempsey and Sandler (1995) presented a five-level model of parental involvement in children's education: parental decision to be involved in their children's education; parents' choice of involvement forms; mechanisms through which parental involvement influences children's learning outcomes; tempering/mediating variables; and child/student outcomes. Because some dimensions are common in the definitions, parental involvement in this thesis is defined as a complex construct, with dimensions as described by Epstein (1995), Hoover-Dempsey and Sandler (1995), and Grolnick and Slowiaczek (1994).

Parental involvement improves student outcomes. A number of studies reported a positive relationship between parental involvement and student achievement, with the relation being stronger at the elementary school level than the secondary school level (Bowen & Lee, 2006; Driessen, Smit, & Sleegers, 2005; Fan & Chen, 2001; Grolnick, Benjet, Kurowski, & Apostoleris, 1997; Grolnick & Slowiaczek, 1994; Izzo, Weissberg, Kasprow, & Fendrich, 1999; Jeynes, 2003, 2005, 2007; Simon, 2001, 2004; Yan & Lin, 2005). The strong relation at the elementary school level is attributed to the elementary school children being relatively easy to influence, less independent in terms of self-regulated learning, and the likelihood of parents spending more time with elementary school children (Jeynes, 2007). Parental involvement in children's schooling is also associated with improvement in school functioning (Izzo, et al., 1999). Corroborating this finding, the OECD (2007a) reported that schools whose principals perceived themselves to be under pressure from parents to maintain high academic standards tended to perform better than schools without such pressure from parents.

The relationship between parental involvement and student outcomes depends on diverse factors. Parents' expectations for their children's educational achievement is a stronger correlate of student achievement than parents' home supervision for their children's studies (Fan & Chen, 2001; Jeynes, 2005, 2007; Yan & Lin, 2005). Parents' high expectations for their children's educational achievement, results in the parents being educationally-oriented, and in the creation of a school-like learning environment at home (Jeynes, 2005; Yan & Lin, 2005). Parental involvement in children's homework through strategies such as modelling, reinforcement, and instruction to foster self-efficacy, self-regulatory skills, and positive attitudes enhances children's success in schools (Hoover-Dempsey, et al., 2001). Provision of cognitive stimulation to children, avoidance of conflicts with children, and more time spent on children's studies by parents also improve children's educational achievement (Eamon, 2005).

Parental involvement is also affected by parents' capacity to engage in children's learning. Hoover-Dempsey and Sandler (1995) identified the following as the factors of parental involvement: parents' role construction; parents' personal sense of efficacy for helping children; parents' reaction to opportunities and demands by children and schools; parents' curriculum domain specific skills and knowledge; parents' modelling, reinforcement, and instruction; parents' use of developmentally appropriate involvement strategies and the fit between parents' involvement actions and school expectations; and parental demands on time and energy. Parents' demographics (SES, one-parent, and dual-parent), parent-child characteristics (parent attitudes and child difficulty), family context (difficult context and social support), school outreach programs, and communication network between school and parents also influence parental involvement in children's education (Desimone, 1999; Grolnick, et al., 1997; Jeynes, 2003, 2007; Ritblatt, Beatty, Cronan, & Ochoa, 2002; Simon, 2001, 2004).

In summary, parental involvement is an effective and powerful family educational resource to improve children's success in schools. The effectiveness of parental involvement and its influence on children's education can be increased by promoting the factors that enhance both parental involvement and their influences. Therefore, the incorporation of parental involvement as a variable in a national educational assessment model will assist parents, teachers, principals, educational policy-makers, and other stakeholders in education to make informed decisions and interventions regarding parental involvement.

3.4.3.5 Educational Leadership

Researchers in educational leadership are supportive of the common belief that educational leaders play a significant role in determining the course of school effectiveness and student outcomes (Hallinger & Heck, 1996; Leithwood & Jantzi, 1990). However, because of the difference in conceptual and operational definitions of educational leadership, researchers have various conflicting conclusions about its relation to student achievement (Hallinger, 1994; Hallinger & Heck, 1998; Witziers, Bosker, & Kruger, 2003).

Two competing models of educational leadership are widely reported in the literature: instructional leadership (Hallinger, 1989, 2003; Scheerens, et al., 2003) and transformational leadership (Hallinger, 2003; Leithwood & Jantzi, 1990, 2005). Hallinger (1989, 2005) conceptualized instructional leadership as having the following three broad dimensions: defining the school mission; managing the instructional program; and promoting the school learning climate. Marks and Printy (2003) summarised the downside of traditional instructional leadership as being paternalistic, archaic and dependent on docile followers which contradict the idea of teacher professionalization (e.g., enabling teachers to play informed and active roles in schooling). However, when teacher commitment, engagement, and competence are low, traditional instructional leadership is seen as an effective leadership model insomuch as it aids in monitoring the teachers (Marks & Printy, 2003). In their review of transformational school leadership research published between 1996 to 2005, Leithwood and Jantzi (2005) identified the following three broad clusters of transformational leadership: setting direction (setting shared vision, group goals, high performance expectations); helping people (providing individualised support; intellectual stimulation, and modelling key values and practices); and redesigning the organization (building collaborative cultures, creating structures to foster collaboration and building better relationships with parents and community). While the transformational leadership model is reported as more popular than the instructional leadership model among schools (Hallinger, 2003; Leithwood & Jantzi, 2005), the combination of the two is reported to have a greater positive effect on school effectiveness as they complement each other (Marks & Printy, 2003). Also, Scheerens et al. (2003, pp. 264-269) conceptualized educational leadership as a complex construct, with ‘general leadership’ and ‘instructional leadership’ as the two underlying dimensions. Scheerens et al. (2003) noted that their concept of educational leadership subsumes other variants in the educational leadership debate and that the concept is strongly related to school effectiveness.

Drawing on Pitner’s (1988) work, Hallinger and Heck (1998) used the following leadership models to review the literature published between 1980 and 1995 on the relation between educational leadership roles and student outcomes: the direct-effects model; the antecedent-effects model; the mediated-effects model; the reciprocal-effects model; and the moderated-effects model. These approaches of studying the association between educational leadership and student outcomes indicate the difficulty of disentangling the pervasive influences of educational leadership in determining the course of school effectiveness. Given such complexities, researchers highlight the need to use these different approaches to study the relation between educational leadership and student outcomes (Hallinger & Heck, 1996, 1998; Maeyer, Rymenans, Petegem, Berg, & Rijlaarsdam, 2007; Marks & Printy, 2003; Witziers, et al., 2003). To illustrate this point, eight case studies of successful principals are summarized in Table 3.13. Table 3.13 depicts states, number of schools involved in the case studies, countries, the researchers who conducted the case studies, and leadership behaviours.

Table 3. 13. Leadership Behaviours of Eight Successful Principals

State (Schools), Country, and Source	Leadership Behaviours
Virginia (12), USA , (Crum & Sherman, 2008)	Developing personnel and facilitating relationship; Sharing responsibility through delegation and team empowerment; Recognizing accountability; Establishing communication and rapport; Facilitating instruction; Managing change

Tasmania (5) and Victoria (9), Australia, (Gurr, Drysdale, & Mulford, 2005)	Understanding school contexts and situations; Principal's values and beliefs; Providing individual support; Building individual capacity; Building school capacity; Developing shared vision and direction; Focussing on school outcomes; Monitoring; Transformation; Recognizing personal characteristics; Styles of leadership; Providing leadership support in teaching and learning
England (10), Britain, (Day, 2005)	Emphasizing moral purpose and social justice; High organizational expectation for learning outcomes; Developing identity, trust, and passionate commitment
Western New York (7), USA, (Jacobson, Johnson, Ylimaki, & Giles, 2005)	Setting direction; Developing people; Redesigning organization; Establishing clear accountability principles; Developing caring principles; Developing learning principles
North Denmark (2), Denmark, (Moos, Krejsler, Kofod, & Jensen, 2005)	Using clear communication strategies; Focusing on student learning; Creating democratic conditions; Sharing leadership responsibilities; Providing feedback and appraisals
Norway (12), Norway, (Moller et al., 2005)	Adopting learning-centred approach; Adopting team-centred leadership; Using skilful balancing actions/negotiations
Sweden (3), Sweden, (Hoog, Johansson, & Olofsson, 2005)	Using school culture and structure as foundations for new interventions
Shanghai (2), China, (Wong, 2005)	Developing positive expectation; Developing opportunities for success; Providing positive feedback; Developing self-evaluation; Using school strengths as foundations for new interventions; Respecting individual needs

Table 3.13 shows that while leadership behaviours differ among countries, the behaviours comprise the combined components of instructional and transformational leadership. The differences in leadership behaviours among countries indicate the influences of the school contexts (e.g., school culture and climate) on leadership behaviours and the difficulty of standardizing leadership behaviours across schools.

Notwithstanding the complexities of studying the leadership effects on student outcomes, researchers have reported a consistent relationship between leadership and student outcomes. The meta-analytic study of 37 research papers published between 1986 and 1996 on the relationship between educational leadership and student achievement showed that they are correlated (Witziers, et al., 2003). In their review of the research literature published during the period from 1980 to 1995 on the relationship between principal leadership and student achievement, Hallinger and Heck (1996, 1998) reported that school principals exercise a measurable, but mostly indirect influence on school effectiveness and student achievement. Hallinger (2005) reiterated the same finding in his recent study of the empirical evidences about the effects of instructional leadership on student outcomes. Also, in concert with Hallinger (2005), Leithwood and Jantzi (2005) reported that school leadership has significant, but indirect effects on student achievement and student engagement in school.

In summary, educational leadership is widely accepted as an influential factor that determines school effectiveness, though its effects on student outcomes are often assessed indirectly. With educational leadership used as a variable in a national educational assessment model, information about the relationship between leadership behaviours and student outcomes could be processed from the national education assessment data for use in developing relevant educational interventions.

3.4.3.6 School Resources

The relationship between school resources and student achievement has been widely studied by researchers since Coleman et al.'s (1966) classic report *Equality of Educational Opportunity*. The research in the area of school resources and student outcomes has identified teacher ability, teacher education, teacher experience, teacher salary, teacher-pupil ratio, per-pupil expenditure, and school size as school resources which correlate with student outcomes (Greenwald, et al., 1996; Hanushek, 1997; Hedges, et al., 1994; Marzano, 2001; Scheerens, 2000; Wayne & Youngs, 2003; Wenglinsky, 2002). Barber and Mourshed (2007, p. 16), in their study of the world's best performing schools, authoritatively noted that "the quality of an education system cannot exceed the quality of its teachers". This statement underscores the superordinate nature of teacher quality over all other educational resources.

Coleman et al. (1966) reported that 10% of the variance in student achievement was explained by school resources. In general, Mullis et al. (2008) reported that students in the schools that were unaffected by the lack of resources had higher average achievement than students in the schools that were affected by the lack of resources.

Scheerens, Glass, and Thomas (2003) assert that school resources are malleable, indicating that the resources can be developed through appropriate policy interventions. Therefore, information about how resources are used by, and distributed to, schools has to be available. One of the ways to get such information is to use the measures of school resources as variables in a national educational assessment model. Such information would help policy-makers and other stakeholders design interventions for school resources.

3.4.4 Educational Effectiveness Factors at Context-Level

Researchers in educational effectiveness have identified context-level educational effectiveness factors as national educational policy, evaluation of educational policy, national education environments, descriptive characteristics of school and its environment, and external achievement stimuli.

Creemers and Kyriakides (2008) identified three overarching context-level educational effectiveness factors. The first overarching factor is a national or regional policy for education. This factor refers to the national educational policy in relation to teaching and learning. On the teaching front, a national policy on school timetables, long-term and the short-term planning, policies on absenteeism and drop-out, and standards for teaching are emphasised. With regard to learning, the national policy on school curriculum is emphasized. The second overarching factor refers to the evaluation of the educational policy. This factor emphasises the need for an ongoing evaluation mechanism for evaluating the national education policy in terms of its effectiveness in improving teaching and learning. The third overarching factor is the overall national education environment. This factor relates to the support (e.g., financial, strategies, advice) provided by different stakeholders to schools, and the expectations (e.g., achievement pressure) from the stakeholders on schools about learning and student outcomes.

Scheerens (1992) identified the following two broad educational effectiveness factors: descriptive characteristics of the school and school environment, and external stimuli to make schooling more effective. The first category is related to the composition of the school population (e.g., disadvantaged pupils, ethnic minorities), the denomination of the school (e.g., private school, public school, boarding school, day school), and the geographical setting of the school (e.g., rural school, inner city school, suburban school). The second category refers to reward structures (e.g., privatization, deregulation, output financing), assessment-centred government policy, parental involvement, demands made by educational consumers, and cultural values (e.g., societal perspectives on the quality of education and the teaching profession).

The value of context-level factors identified by Creemers and Kyriakides (2008) and Scheerens (1992) lies in their potential use as cues and references for studying national educational policy guidelines and directives and in identifying national-level educational effectiveness factors. Therefore, the context-level factors when used in a national educational assessment model have the potential to generate information about the influence of national education policies on student outcomes. Such information should help stakeholders in Bhutan’s education system to make data-driven policy decisions and interventions, and more wisely develop evaluation plans and strategies.

3.5 The National Educational Assessment Model with Educational Effectiveness Factors

The preceding sections identified the key educational effectiveness variables for all the levels of the proposed national educational assessment model. This section fleshes out the proposed national educational assessment model with the key educational effectiveness variables. The section also describes how the proposed model will be validated by involving diverse stakeholders in the Bhutanese education system.

3.5.1 Fleshing Out the National Educational Assessment Model with Educational Effectiveness Factors

The proposed national educational assessment model can now be fleshed out with educational effectiveness factors relevant to its levels. First, the model is a multi-level structure, with the structure consisting of context-, school-, classroom-, and student-levels. Second, the model uses the input-process-output paradigm. Third, as discussed, the model is underpinned by the assumptions of the Integrated School Effectiveness Model and the Dynamic School Effectiveness Model. Fourth, context-, school-, classroom-, and student-levels of the model are characterized in terms of the educational effectiveness factors described in the literature of educational effectiveness research and large-scale educational assessments. Figure 3.4 shows the proposed model with the educational effectiveness factors. Tests in this model refer to high-stakes tests reviewed in Chapter 2.

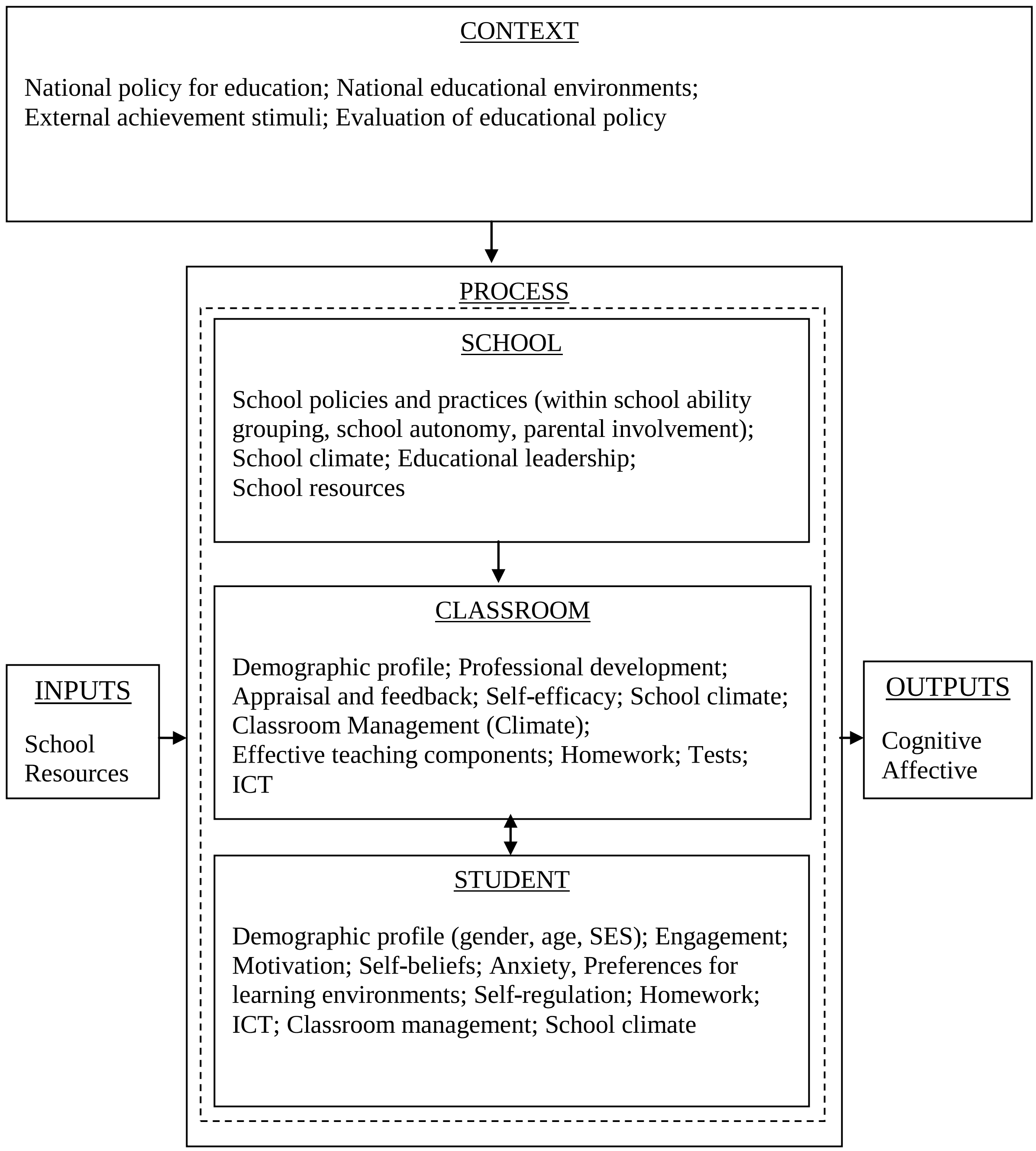


Figure 3. 4. National Educational Assessment Model Expanded

3.5.2 Validating the Proposed Model

As discussed in Section 2.6.1 of Chapter 2, a consensual approach to developing an assessment model gives the model wider ownership as well as strengthening its validity. To accord these attributes to the proposed model, the model was presented to a panel of key stakeholders in the Bhutanese education system. The stakeholders comprised representatives from the Department of

School Education, the Education Monitoring and Support Service Division, the Department of Curriculum and Professional Support Services, the Policy and Planning Division, the Bhutan Board of Examinations, and the Human Resource Division of the Ministry of Education. The panel, held for two hours, was chaired by the Secretary of the Ministry of Education.

The overall view of the panel was that the proposed model would be of great use to the Ministry of Education in evaluating the quality of the Bhutanese education system. However, as expected, some panel members pointed out that the proposed model appeared ambitious in its scope, but noted that the model would be able to provide a comprehensive knowledge about the quality of the Bhutanese education system if all of its aspects could be implemented. None of the panel members differed in their perceptions of the relevance of the model and its various educational effectiveness factors, indicating its relevance to the Bhutanese education system.

Furthermore, the view of the panel members that the model would be useful in gauging the quality of the Bhutanese education system is indicative of the model's potential to drive educational improvement programmes in Bhutan. All this indicated that the model was viewed as valid and accepted by the key stakeholders in the Bhutanese education system.

In summary, the proposed national educational assessment model is developed by using the Integrated School Effectiveness Model and the Dynamic Model of School Effectiveness as starting points. The proposed model is multi-level in nature, is underpinned by the same assumptions on which the Integrated School Effectiveness Model and the Dynamic Model of School Effectiveness are based, and it has educational effectiveness factors commonly described in the literature on educational effectiveness research. Table 3.14 provides a comparison of the educational effectiveness factors of the Integrated School Effectiveness Model, the Dynamic School Effectiveness Model, and the National Educational Assessment Model.

Table 3. 14. Comparative List of Educational Effectiveness Factors Identified by Different Models

Integrated School Effectiveness Model	Dynamic School Effectiveness Model	National Educational Assessment Model
<u>Context-Level Factors</u> • Achievement stimuli from higher administrative levels • Development of educational consumerism; • Co-variables like school size, student-body	<u>Context-Level Factors</u> • National/regional policy for education • Evaluation of policy • Educational environment <u>School-Level Factors</u>	<u>Context-Level Factors</u> • National policy for education • Evaluation mechanism for educational policy • National educational environment (e.g., educational consumerism, school category) • External achievement stimuli

composition, school category, and urban/rural distinctions <u>School-Level Factors</u> • Achievement-oriented policy • Educational leadership • Consensus, co-operative planning of teachers • Quality of school curricula in terms of content covered, and formal structure • Pressure for achievement; • Recruitment of qualified staff • Financial and material Characteristics of the school • Orderly atmosphere • Evaluative potential <u>Classroom-Level Factors</u> • Time on task • Structured teaching • Opportunity to learn • High expectations of pupils' progress • Evaluation and monitoring of pupils' progress • Reinforcement	• School policy and evaluation of school policy <u>Classroom-Level Factors</u> • Quality of teaching (orientation, structuring, modelling, application, questioning, assessment, management of time, classroom as a learning environment) <u>Student-Level Factors</u> • Aptitude • Perseverance • Time on task • Opportunity to learn • SES • Gender • Ethnicity • Personality traits • Expectations • Thinking style • Subject motivation	<u>School-Level Factors</u> • School policies and practices • School climate • Educational leadership • School resources <u>Classroom-Level Factors</u> • Demographic profile • Professional development • Appraisal and feedback • Self-efficacy • School climate • Classroom Management (Climate) • Effective teaching components • Homework • Tests • ICT <u>Student-Level Factors</u> • Gender • Age • SES • Motivation • Self-efficacy • Self-regulation • Learning preferences • Homework • ICT • Classroom management • School climate • Engagement

As shown in Table 3.14, the models have similar educational effectiveness factors. All three models use teaching components as educational effectiveness factors at the classroom level. The proposed national education assessment model differs from the other two models by its inclusion of ICT as an educational effectiveness factor.

3.6 Chapter Summary

Educational effectiveness is a field of study that seeks to explain why some schools and teachers are more effective than others, with effectiveness being the extent to which educational processes lead to attainment of educational goals. Over the years, educational effectiveness study has evolved into different traditions, with later traditions claiming to be more comprehensive than their predecessors. The evolving five traditions of educational effectiveness detailed earlier in this chapter have contributed to the skeletal system-wide educational assessment model for Bhutan (see Figure 3.4).

The national educational assessment model for Bhutan is then fleshed out with contextual variables and developed into a complete national educational assessment model. All contextual variables in the national educational assessment model are related to student outcomes, and they have the potential to contribute to quality education by assisting stakeholders in planning and implementing school improvement programmes. In addition, the national educational assessment model is accepted by the Ministry of the Education and other key stakeholders in the Bhutanese education system.

3.7 Chapter Conclusion

This chapter examined critically various traditions of educational effectiveness research. Drawing on the strengths and weaknesses of individual traditions, a skeletal model of a system-wide educational assessment model has been developed as presented in Figure 3.4. With the contextual variables of NAEP, TIMSS, and PISA as the starting points, various educational effectiveness variables were examined. The factors were then used in fleshing out the skeletal national educational assessment model, resulting in a complete national educational assessment model. Further, the complete national educational assessment model has been validated by presenting it to a panel. The panel endorsed the model.

Chapter 4

METHOD

4.1 Introduction

This chapter examines critically a range of research designs to identify research methods for collecting data on the educational effectiveness factors assigned to different levels of the national educational assessment model presented in Chapter 3. The chapter also presents a case for the most appropriate research methods, the most relevant research instruments, and the most appropriate sampling designs for the study. In addition, the field administrations of the research instruments, including validity issues, and the preparation of data for analyses in the later chapters are described in this chapter. The chapter also describes the procedures followed in calibrating the mathematics test items, linking the Mathematics test to PISA 2003, and developing a mathematics proficiency scale.

The first section of the chapter revisits the research aims and outcomes presented in Chapter 1, the second section presents the research designs, the third and the fourth sections present the sampling design, and the fifth section presents the data collection instruments used in the study. The fifth section also describes the validation of the data collections instruments. The final sections present the chapter summary and conclusion.

4.2 Revisiting the Research Aim and Outcomes

Chapter 1 highlighted the current need for the Kingdom of Bhutan to have a national educational assessment programme capable of providing a range of data-based information about the effectiveness of its education system and guiding stakeholders in making research-based policy decisions and school improvement programmes. In addition, the knowledge from the national educational assessment programme should enable the stakeholders to bridge the gap between the policy directives and the field realities. These broad functions of the national educational assessment programme are further specified into 10 specific outcomes, as listed in Chapter 1. Drawing on these 10 outcomes, a national education assessment model has been proposed in Chapter 3. The proposed model has four levels, namely, context, school, classroom, and student levels. Depending on its aim and outcomes and the educational effectiveness factors identified in its proposed model, the study needs a different research design for collecting data.

4.3 Research Designs

A research design comprises plans, paradigms, and methods used by researchers to address research questions, problems, and hypotheses (Borg & Gall, 1989; Creswell, 2005; Kerlinger, 1989;

Krathwohl, 2009; Lincoln & Guba, 1985). The selection of a research design is largely influenced by a researcher’s choice of paradigms, which further depends upon the research questions, problems, and hypotheses. This section examines critically some common research paradigms by focussing on their relevance to the aim and outcomes of this study as well as focussing on their strengths and weaknesses.

4.3.1 Research Paradigms

Guba and Lincoln (1994, p. 105) defined a research paradigm as “the basic belief system or worldview that guides the investigator, not only in choices of method but in ontologically and epistemologically fundamental ways”. The following five paradigms are commonly identified in the research literature: positivism (Cook, 1985; Guba, 1990; Guba & Lincoln, 1994; Maykut & Morehouse, 1994; D. C. Phillips & Burbules, 2000); postpositivism (D. C. Phillips, 1983, 1994; D. C. Phillips & Burbules, 2000); critical theory (Kincheloe & McLaren, 2003; Popkewitz, 1990); constructivism (Guba, 1990; Guba & Lincoln, 1994; Lincoln & Guba, 1985); and pragmatism (Creswell, 2003; R. B. Johnson & Onwuegbuzie, 2004; Onwuegbuzie, 2002; Reichardt & Rallis, 1994; Tashakkori & Teddlie, 1998). The above five paradigms differ in their ontology, epistemology, and methodology. They also differ in the degree to which they have been accepted by researchers.

4.3.1.1 Positivism

Ontologically, positivism is characterised as realist belief in the existence of an independent and understandable reality and is committed to knowing the true nature and the functioning of such reality by studying its parts to understand its whole, with the goal of predicting and controlling its natural phenomena (Cook, 1985; Guba, 1990; Guba & Lincoln, 1994; Maykut & Morehouse, 1994; D. C. Phillips & Burbules, 2000). Positivist knowledge claims about reality, once known, are viewed as “time-and context-free generalizations, some of which take cause-effect laws” (Guba, 1990, p. 20; Guba & Lincoln, 1994, p. 109). A strong critique of positivism has been that the time- and context-free property of a knowledge claim is not relevant to social science; simply because variability is bound to exist across individuals, ethnic groups, cultures, and nationalities of people related to the knowledge claim (Borg & Gall, 1989; Lincoln & Guba, 1985; Maykut & Morehouse, 1994). Epistemologically, positivism is characterised as “dualist” because it views researchers and the focus of their research as independent entities (Cook, 1985; D. C. Phillips, 1994; D. C. Phillips & Burbules, 2000). Positivism is also characterised as “objectivist” because it requires research interest to be independent of researchers’ values (or vice-versa), and facts from research data to be independent of theory. The critics of positivism persuasively argued that facts are both theory- and value-laden because facts are only meaningful within a theoretical framework, that theories are themselves value statements, and that evidence should not underdetermine theories (Guba & Lincoln, 1994; Lincoln & Guba, 1985; D. C. Phillips & Burbules, 2000). Methodologically, positivism is associated with the use of empirical tests to verify research outcomes and the use of research designs to control for unsolicited noise, confounding factors, and potential threads capable of unduly influencing the research outcomes (D. C. Phillips, 1994; D. C. Phillips & Burbules, 2000). The critics of positivism are of the view that not all educational research can be subjected to empirical tests to verify or falsify their findings or claims (Guba, 1990; Guba & Lincoln, 1994). Various criticisms of positivism have led to the emergence of postpositivism.

4.3.1.2 Postpositivism

Ontologically, advocates of postpositivism believe in the existence of reality that can only be partially understood and consider it to be fallible in relation to the weight of evidence (D. C. Phillips & Burbules, 2000). Guba and Lincoln (1994, p. 110) noted convincingly that the postpositivists ontological notion of reality acknowledges the "flawed human intellectual mechanisms and fundamentally intractable nature of phenomena" which they claim is one of the fundamental tenets of constructivism. Epistemologically, postpositivism emphasises objectivity without adhering to a positivist idea of dualism (1994; D. C. Phillips, 1994; D. C. Phillips & Burbules, 2000). Contrary to positivism, postpositivism believes that observation is theory- and value-laden, that theory is underdetermined by facts, and that a nascent knowledge claim should be scrutinized by communities of knowledge producers for potential bias in research design (D. C. Phillips, 1994). However, postpositivism differentiates values as external (epistemically irrelevant) and internal (epistemically relevant). Phillips and Burbules (2000) identified epistemically irrelevant values as political, religious, power, social, economic, and cultural influences on researchers. On the other hand, Phillips and Burbles (2000, p. 54) defined epistemically relevant values as researchers' "dedication to the pursuit of truth, openness to counter evidence, receptiveness to criticism, accuracy of measurements and observations, honesty and openness in reporting results, and the like". Postpositivist recognition of the criticism of the positivist value neutrality of scientific inquiry reflects epistemically irrelevant values, not epistemically relevant values (D. C. Phillips & Burbules, 2000). Phillips and Burbules (2000, p. 54) aptly noted that epistemically relevant values are "constitutive of scientific enquiry, that is, without them scientific enquiry loses its point". Methodologically, the rigid experimental and manipulative strictures in positivism are relaxed in postpositivism. Phillips and Burbules (2000, pp. 86-97) noted that "postpositivist approach to research is based on seeking appropriate and adequate warrants for conclusions, on hewing to standards of truth and falsity that subject hypotheses (of whatever type) to test and thus potential disconfirmation, and on being open-minded about criticism". This statement implies that postpositivism relates more to quantitative than to qualitative research methods. A strong critique of postpositivism has been its emphasis on the need to dissociate epistemically irrelevant values of researchers from their knowledge claims (Kincheloe & McLaren, 2003) unlike the critical science paradigm that accounts for epistemically irrelevant values.

4.3.1.3 Critical Science

A critical science paradigm is concerned with social, political, cultural, and economic dimensions of social systems (Kincheloe & McLaren, 2003; Popkewitz, 1990). This paradigm is underwritten by critical theory which includes a range of other theories (e.g., feminism, poststructuralism, neo-Marxian theory) associated with social issues (Guba, 1990; Guba & Lincoln, 1994; Kincheloe & McLaren, 2003; Popkewitz, 1990). Kincheloe and McLaren (2003) acknowledged the difficulty of presenting a common definition of critical science for the following reasons: critical science keeps changing and evolving, critical science is based on different critical theories, and critical theorists often differ in their views about social issues. However, researchers have converged on recognising ontological, epistemological, and methodological characteristics of critical science (Guba, 1990; Guba & Lincoln, 1994; Howe, 1992; Kincheloe & McLaren, 2003; Popkewitz, 1990). Ontologically, critical science believes in the existence of reality that is only partially apprehendable. Epistemologically, critical science views knowledge claims as subjective because knowledge claims are subject to influence by the values and contexts associated with researchers. Methodologically, critical theorists use dialogic, hermeneutic, and transformative approaches to generating knowledge claims. Guba (1990, p. 24) convincingly noted that the ontology and epistemology of critical science present a “logical disjunction” because of the former’s alignment with the ontology of the postpositivists. This has led to the emergence of constructivism.

4.3.1.4 Constructivism

Ontologically, constructivism is associated with relativism because constructivists believe that realities are comprehensible in the form of numerous intangible mental constructions of the inquirer and the inquired as they interact (Guba, 1990; Guba & Lincoln, 1994; Lincoln & Guba, 1985). Constructivists also believe that the constructed realities evolve, depending upon the time and the environment (social, cultural, economic, and political) in which the inquirer and the inquired are situated (Guba, 1990; Guba & Lincoln, 1994; Lincoln & Guba, 1985). Epistemologically, constructivists are subjectivists. Constructivists believe that knowledge claims are both theory- and value-laden, that the inquirer and the subject of inquiry are interactively linked, and that research findings are created as the inquiry progresses (Guba, 1990; Guba & Lincoln, 1994; Lincoln & Guba, 1985). Methodologically, constructivists use hermeneutic and dialectic methods to generate constructions on which there is substantial consensus (Guba, 1990; Guba & Lincoln, 1994; Lincoln & Guba, 1985). The differences among positivism, postpositivism, critical science, and constructivism in their scope for meeting the needs of researchers fostered a pluralist paradigm that has the flexibility of consolidating the differences and offering researchers a broader scope for choosing research methods appropriate to their interests.

4.3.1.5 Pragmatism

Pragmatism has emerged as a research paradigm advocating the need to take the pluralist approach to research (Creswell, 2003; R. B. Johnson & Onwuegbuzie, 2004; Onwuegbuzie, 2002; Reichardt & Rallis, 1994; Tashakkori & Teddlie, 1998). The pragmatist ontology is the combination of ontological aspects of postpositivism and constructivism. Pragmatists believe that reality is partially knowable and that it exists in multiple forms (R. B. Johnson & Onwuegbuzie, 2004). Pragmatists also view actions, situations, and consequences as factors influencing knowledge claims, not antecedent conditions (Creswell, 2003; R. B. Johnson & Onwuegbuzie, 2004; Tashakkori & Teddlie, 1998). Epistemologically, pragmatists reject dualisms and reductionism. Pragmatists believe that researchers function within certain contexts, and thus research findings bear the values of the contexts in which researchers work (Creswell, 2003; R. B. Johnson & Onwuegbuzie, 2004; Tashakkori & Teddlie, 1998). In fact, pragmatists believe that research is both theory- and value-laden. Methodologically, pragmatists can choose their methods from a range of qualitative and quantitative methods. Johnson and Onwuegbuzie (2004, p. 18) plausibly noted that pragmatists endorse “eclecticism and pluralism (e.g., different, even conflicting theories and perspectives can be useful; observation, experience, and experiments are all useful ways to gain an understanding of people and the world)”.

4.3.1.6 Choice of a Paradigm

Research problems determine the choice of a paradigm for a research study. It has been shown in the proposed national educational assessment model, presented in Chapter 3, that a holistic evaluation of the Bhutanese education system can be made by using cognitive tests and educational effectiveness factors. The underlying assumptions of cognitive tests are that it is possible to obtain a close approximation of a student’s accurate knowledge, that this accurate account of knowledge can be described in terms of norms and criteria, and that the true knowledge can be obtained through reliable tests (Aiken, 1997; Cohen & Swerdlik, 1999). These assumptions imply that the proposed model seeks to generate a knowledge claim from cognitive tests and that the knowledge claim is a close approximation of students’ true score. The proposed national educational assessment model also indicates that the student outcomes, including performance on cognitive tests, depend on other educational effectiveness factors with the assumption that the factors are measurable through abstraction by operational language. In addition, the proposed model seeks to generalize findings from a sample population to the target population. These aspects of the proposed model fit well with postpositivism. The model assumes that the educational effectiveness factors are multi-level, that the higher levels provide a supportive environment for the lower levels, that the relationship among factors at different levels is constant relative to time, and that the effectiveness factors operate differentially across levels in line with contingency theories. These assumptions align well with the attributes of critical science and constructivism. The fit of the proposed model with postpositivism, critical science, and constructivism indicates that the proposed model needs a pluralist research paradigm. As described earlier, pragmatism is a pluralist paradigm; therefore, pragmatism has been chosen as the most suitable paradigm for the proposed model.

4.3.2 Research Methods

The chosen research paradigm guides the choice of a research method for the study. Generally, researchers associate positivism and postpositivism with quantitative methods, and critical science and constructivism with qualitative methods. However, the observation that quantitative and qualitative methods are difficult to pursue independently in an absolute sense (R. B. Johnson & Onwuegbuzie, 2004; Tashakkori & Teddlie, 1998) favours a methods continuum with quantitative and qualitative methods at the extreme ends (Borg & Gall, 1989; Creswell, 2003, 2005; Krathwohl, 2009; Onwuegbuzie, 2002; Tashakkori & Teddlie, 1998), resulting in mixed methods. Mixed methods research is generally associated with the pragmatic paradigm (Creswell, 2003, 2005; R. B. Johnson & Onwuegbuzie, 2004; Kerlinger, 1989; Krathwohl, 2009; Onwuegbuzie, 2002; Tashakkori & Teddlie, 1998). As this study is founded on the pragmatic paradigm, as mentioned earlier, mixed methods becomes a natural choice of method for the study.

A critical examination of the attributes of quantitative and qualitative methods in the light of the salient features of this study shows that a mixed-method approach is appropriate for the study. Corroborating Creswell (2003, 2005), Krathwohl (2009) noted the use of objective inquiry, the collection of numeric data, and the statistical analyses of data by researchers to address their research problems as salient characteristics of a quantitative method. According to Krathwohl (2009), a quantitative method is appropriate for addressing a research problem that has substantive literature information and predefined variables, which is the case with this study because the development and choice of educational effectiveness factors for the proposed national educational assessment model is guided by substantive literature. A quantitative method also uses measures or numbers to describe, explain, and explore research problems (Krathwohl, 2009). As described in the previous chapters, relationships between variables of educational effectiveness factors and student outcomes are commonly described in terms of measures or numbers, making incorporation of a quantitative method appropriate for this study. Krathwohl (2009) also noted that a quantitative method is appropriate when researchers study their problems within the contexts of theories or hypotheses. Relating this criterion to the variables described in Chapter 3, the emerging hypothesis has been that the variables are positively related to student outcomes. However, it needs to be explored in depth whether a similar relationship exists in Bhutanese schools. Krathwohl's (2009) guidelines on using quantitative methods and their application by NAEP, TIMSS, and PISA show that similar paradigms and methods are suitable for this study.

As emphasised by many writers on research methodology, the use of a single method in a study has lots of limitations (Creswell, 2003, 2005; Kerlinger, 1989; Krathwohl, 2009). A quantitative method is dominated by numerical data that limit the research scope to generating statistical information by using a priori close-ended questions. The narrow and specific quantitative close-ended questions require researchers to envisage and know all information relevant to their research problems. Such questions risk constraining research participants from expressing their views and experiences because the quantitative close-ended questions do not stimulate them, though they may be related to researchers' interests. This kind of risk is even greater when the phenomenon under study is related to contingency theories (Donaldson, 2001). The proposed national educational assessment model assumes that the effectiveness factors operate differentially across levels in line with contingency theories, emphasizing the context specificity of educational effectiveness factors. Qualitative open-ended questions have their relevance to this study where context specific responses from the research subjects are required. The context level of the proposed model needs a qualitative method because data on the educational factors identified at this level have to be collected through interactive communication with the data sources.

In summary, this study is underpinned by pragmatism as its paradigm and mixed methods as its research method.

4.3.2.1 Cross-Sectional Survey

Among various research design options that mixed methods studies have, this study uses cross-sectional survey (Babbie, 1990; Fowler, 2009; Rea & Parker, 2005) and focus-group interviews (Morgan, 1988) designs concurrently. Tashakkori and Teddlie (1998) called such a design concurrent mixed methods.

Surveys are commonly used as data collection tools by researchers who seek to describe, explain, and explore certain traits or attributes of a population by using sample- or census-based data of the population (Babbie, 1990; Fowler, 2009; Rea & Parker, 2005). With descriptive, explanatory, and exploratory information about a sample, researchers generalize their findings from a sample to a population. As the purpose of this thesis is to provide stakeholders in the Bhutanese education system with descriptive, explanatory, and exploratory information about their education system, a cross-sectional sample survey is deemed most appropriate for collecting data for this study.

In addition to its close fit with the purpose of this study, a sample survey also has other advantages (Babbie, 1990; Fowler, 2009; Rea & Parker, 2005) relevant to the study. First, a sample survey allows a researcher to generalize about a whole population by using data from a small portion of that population. This saves cost and time of a researcher, which is not possible if data were collected from the whole population. This aspect of a sample survey is relevant to this study in that the timely completion of the study will assist the Ministry of Education in its ongoing endeavours to improve the quality of education as well as enabling this study to be completed with limited financial resources. Second, researchers can use data from their sample surveys to develop a database for secondary research. This directly relates to one of the outcomes of this study, which is to develop Bhutan's national educational assessment database for secondary research. Third, robust findings from the sample survey can be used for future reference in studying educational changes over time. This is relevant to this study because education is a dynamic process, as assumed by the proposed national educational assessment model that reflects the need to study educational changes over time.

Between longitudinal and cross-sectional survey designs, the cross-sectional survey is used in this study because the data for the study were collected only at a point in time, that is, while the thesis was being written. However, cross-sectional surveys can be used to approximate a longitudinal survey (Babbie, 1990). Some of the devices for incorporating the features of a longitudinal survey in a cross-sectional survey are: the use of questionnaire items related to process; the use of age or cohort comparisons within a cross-sectional survey; and the use of logical progression to show process over time (Babbie, 1990). These flexibilities of a cross-sectional survey were incorporated in the questionnaires to collect data on the information related to over-time changes in the Bhutanese education system. Specifically, the cross-sectional survey design is aimed at collecting data for the questions related to different levels of the proposed national educational assessment model as follows:

Student Level (StL)

StL 1 How well do gender, age, and SES relate to student achievement?
StL 2 How well does motivation relate to student achievement?
StL 3 How well do self-beliefs relate to student achievement?
StL 4 How well do meta-cognitive skills relate to student achievement?
StL 5 What are students learning preferences?
StL 6 How well does ICT relate to student achievement?
StL 7 What do students think of their classroom and school climate?
StL 8 How well does homework relate to student achievement?

StL 9 How well do students rate on the higher-order thinking skills scale?

StL 10 Are Bhutanese students engaged in their schools?

StL 11 Are Bhutanese students prepared for the challenges of the future?

StL 12 Do Bhutanese students have mathematical knowledge and skills to adapt to rapid societal and technological change?

StL 13 How does Bhutanese student achievement vary across school, district, regional, national levels?

StL 14 How does Bhutanese Grade 10 student achievement compare with international students?

Classroom Level (ClL)

ClL 1 How well do gender, age, educational qualification, and teaching experience of teachers relate to student achievement?

ClL 2 How well do Bhutanese teachers' demographic profiles compare with those of teachers across the OECD countries?

ClL 3 How well does teacher professional collaboration and development relate to student achievement?

ClL 4 How well do teacher appraisal and feedback relate to student achievement?

ClL 5 How well does classroom management relate to student achievement?

ClL 6 How well does the teachers' view of their school climate relate to student achievement?

ClL 7 How well do teachers' beliefs and practices of teaching strategies relate to student achievement?

ClL 8 How well does teachers' self-efficacy relate to student achievement?

ClL 9 How well do teachers' usage of homework, ICT, calculators, and tests relate to student achievement?

ClL 10 How do teachers perceive their student engagement?

School Level (ScL)

ScL 1 How well do school policies and practices of student admittance, ability grouping, assessment and accountability, and parental involvement relate to student achievement?

ScL 2 What is the state of school autonomy in Bhutan?

ScL 3 How well do school resources relate to student achievement?

ScL 4 How well do schools rate on the school climate scale?

ScL 5 How well do schools rate on an instructional leadership model?

ScL 6 What is the state of access and equity in school resources?

Inputs Level (InL)

InL 1 What educational resources are available in schools?

Outputs Level (OuL)

OuL 1 What is the profile of Bhutanese Grade 10 students' mathematical knowledge and skills at the end of their basic education?

OuL 2 What are the affective responses of Bhutanese students to their basic education?

The next section describes focus group interviews and their use in collecting information at the context level of the proposed national educational assessment model.

4.3.2.2 Focus Group Interview

Among various data collection instruments available for constructivism, a focus group interview is used in this study because it fits with the evolutionary nature of policies and the emerging attitudes, perspectives, and cognitions about educational effectiveness factors identified at the context level of the proposed national educational assessment model. Morgan (1988, p. 12) noted that a salient feature of focus groups is the "explicit use of the group interaction to produce data and insights that would be less accessible without the interaction found in a group". This hallmark of the focus group aligns well with the nature of the knowledge required at the context level of the national educational assessment model. The context level of the national educational assessment model is related to the national education policies that shape the context in which Bhutanese schools function. In line with the dynamic aspect of educational processes, education policies evolve. Therefore, it is considered appropriate to use focus groups to collect information about the factors identified at the context level of the proposed national educational assessment model by involving people who have been formulating national education policies. This also then makes the purposive sampling method (Krueger & Casey, 2009; Morgan, 1988; Morgan & Scannell, 1998) suitable for selecting participants for the focus group interview as applied in this study.

It is an established practice in Bhutan for a policy to undergo a series of steps similar to those described in the policy cycle presented by Bridgman and Davis (2007). Bridgman and Davis's (2007) policy cycle has the following eight steps: issue identification, policy analysis, policy instruments, consultation, coordination, decision, implementation, and evaluation. Bridgman and Davis (2007) noted that the main persons responsible for processing a potential policy at the various stages of the policy cycle are civil servants, which certainly applies to Bhutan. At the Ministry of Education, the programme officers are responsible for the policy-related tasks. Therefore, the attitudes, perspectives, and knowledge of the programme officers about the various educational effectiveness factors have the potential to provide answers to the context-level questions. In addition, as policies are evolutionary, with the nature of their evolution determined by political, cultural, and social needs, interaction among relevant programme officers on the various educational effectiveness factors is the key to understanding the policy dynamics (Parsons, 1995).

The focus group interview is included in the design to collect information pertinent to the context of the Bhutanese education system, with a special focus on the following points:

Context Level

CoL 1 National education policies on effective learning environments (school climate, teachers' self-efficacy, teachers' performance appraisal, teachers' professional development, school leadership, student well-being, school resources).

CoL 2 External achievement stimuli (parental involvement, choice of schools, school privatization, school autonomy, accountability, and output financing).

CoL 3 Mechanisms for evaluating national education policies (issue identification, policy analysis, policy instruments, consultation, coordination, decision, implementation, and evaluation).

4.4 Sample Design for the Cross-Sectional Survey

A sample design comprises the process of selecting the sample and estimating sample statistics. The quality and quantity of information processed from a cross-sectional survey depends on its sample design. The choice of a sampling design is determined by a number of competing constraints such as the research aim and outcomes, the nature of the population, the availability of resources, the political situation, and the accuracy of the sample estimates (Cochran, 1977; Kish, 1995; K. N. Ross, 1992; Scheaffer, Mendenhall(III), & Ott, 2006). The following sections describe how these factors influenced the development of the sample design for this study.

4.4.1 Research Aim and Outcomes

The aim and outcomes of this study have been presented in Chapter 1. These highlighted the need to compare and differentiate the knowledge generated from data at different levels of aggregation, requiring the cross-sectional sample survey to be more analytical and less descriptive. Consequently, the sample population needs to be precise enough to facilitate data-driven inferences at different levels of aggregation.

4.4.2 Structure of the Bhutanese School Education System

The general structure of the Bhutanese school education system has been described in Chapter 1. Some additional information relevant to choosing a sampling design for this thesis is briefly described as follows.

Schools in Bhutan are categorised into the following four categories: urban, semi-urban, semi-remote, and remote (Policy and Planning Division, 2008, 2009). These categories have been used by the Ministry of Education as the bases for making policy priorities (Policy and Planning Division, 2007, 2008, 2009). Since almost all schools offering Grade 10-level education are government schools, the overall functioning of the schools is directly related to the policy priorities of the Ministry of Education. This relationship implies that schools within a category are homogeneous in terms of receiving professional support and educational resources from the Ministry of Education. In addition, schools within a category tend to enrol students with similar socio-economic status. Further, streaming or classifying students on the basis of race, ethnicity, and other forms of social classification are hardly visible in Bhutan because of the absence of a formal policy on such practices. The presence of school categories and schools with homogeneous characteristics within these categories aligns well with a stratified random sample design. However, schools within a category will have students differing in their learning outcomes, and the same is the case with students within schools. The presence of heterogeneous student outcomes within schools meant that using a cluster sample design, with a group of students from individual schools forming clusters, would decrease the between-school variance, survey costs, and survey administration time. Therefore, schools have to be selected first followed by the selection of clusters of students from each one of the selected schools. The use of this procedure means applying a two-stage cluster sampling design with schools sampled at the first stage and clusters of students sampled at the second stage.

Furthermore, because the national educational assessment model proposed in Chapter 3 is hierarchical, with students nested within classes, classes nested within schools, and schools nested within context, a stratified two-stage cluster sampling design is appropriate for the study. Researchers have noted a number of advantages in using a stratified two-stage cluster sample design over a simple random sample design. Ross (1992) and Scheaffer et al. (2006) noted that a stratified random sample design is cost-effective compared to simple random sample design. Similarly, cluster sample design is more cost-effective than simple random sample design (Kish, 1995). Cochran (1977) noted that a stratified random sample design can yield greater precision in the estimates of population values of the whole population than a simple random sample design, because the former allows stratification of a heterogeneous population into homogeneous sub-populations. Moreover, Scheaffer et al. (2006) pointed out convincingly that researchers can produce estimates of population values of subgroups of the whole population by using a stratified random sample design. Such estimates underpin the research questions of this thesis, as described in Chapter 1. Thus, a stratified two-stage cluster sample design was used in this thesis to sample Grade 10 students in Bhutan. A similar sampling design was also used by TIMSS (IEA, 2004) and PISA (OECD, 2005d, 2009b).

4.4.3 Enabling Environment for Field Research

The political goodwill and favourable logistics for this thesis have been in abundance throughout its writing. This thesis is written at a time when the concern about the quality of the Bhutanese school education system is increasingly emphasized by the Ministry of Education; consequently, the interest of the Ministry in the thesis has been high. However, like most of the survey research, the resources, research objectives, and desired target population also determine the desirable level of sampling accuracy, and hence the choice of a sampling design that has high accuracy and high efficiency: these attributes are achievable with a stratified two-stage cluster sampling design.

4.4.4 Sampling Frames and Units

The registration record of Grade 10 students available with the Bhutan Board of Examinations (BBE) was used as the sampling frame, with students and schools as sampling units in the defined target population.

According to the registration record, 9,213 Grade 10 students from 71 schools were registered with the BBE for the 2009 Bhutan Certificate of Secondary Education Examination. Because the BBE is the only examination board that conducts national-level examinations for Grade 10 students, all Grade 10 students and their schools had to register with it for the Bhutan Certificate of Secondary Education. Therefore, it was possible to sample students and schools from the defined target population with a known probability of selection by using the registration record available at the BBE as the sampling frame. Also the local publication *Annual Education Statistics* (Policy and Planning Division, 2009) was used to get up-to-date information about school categorization and student enrolment numbers in different categories of schools in Bhutan in 2009.

4.4.5 Research Target Populations

Researchers commonly define three types of populations for sampling. These population groups are as follows: the desired target population, the defined target population, and the excluded population (IEA, 2008; OECD, 2005d).

In accordance with the reasons mentioned in Chapter 1, Grade 10 students of Bhutan in the year 2009 were identified as the desired target population for the survey. As the by-products of sampling schools and students, teachers teaching Grade 10 students and principals of the schools offering

Grade 10-level education were considered as part of the population. However, not all schools and students comprising the desired target population could be included in the research because of geographic, demographic, special educational needs, political, socio-economic, and cultural constraints. Therefore, all mainstream schools (i.e., without students with special educational needs) that offered Grade 10-level education and all Grade 10 students who had to attend classes on a regular basis were defined as the target population from which a random sample of schools and students was drawn. The excluded population comprised the difference between the desired and the defined target populations. As exclusions can occur either at the level of a school or within a school, the excluded population included both school-level and within-school exclusions. The following criteria were followed while excluding schools: (a) the excluded school had to be a special education school; (b) the excluded school had only Grade 10 students who were exempted from attending regular classes; and (c) the excluded school was deemed geographically inaccessible during the period of the field research. The within-school exclusions consisted of students enrolled in the mainstream schools who did not have to attend regular classes. Table 4.1 presents the summary of the three groups of the population.

Table 4. 1. Description of the Desired, Defined, and Excluded Populations

School Category	Desired		Defined		Excluded	
	Number of Schools	Number of Students	Number of Schools	Number of Students	Number of Schools	Number of Students
Urban	41	6099	40	6081	1	18
Semi-Urban	22	2383	17	1964	5	419
Semi-Remote	5	385	2	222	3	163
Remote	3	346	1	197	2	149
Total	71	9213	60	8464	11	749

As is apparent from Table 4.1, 11 schools (15%) and 749 students (8%) were excluded from the desired target population. Out of the 11 excluded schools, one was urban, five were semi-urban, three were semi-remote, and two were remote schools. The one excluded urban school and one of the excluded semi-urban schools had only continuing education or private students. In Bhutan, continuing education candidates are mature-age students who seek to upgrade their qualifications without taking regular classes unlike other mainstream students. Private candidates are students who did not qualify for promotion to Grade 11 in the previous year, and thus were registered for the

Bhutan Certificate of Secondary Education Examination in the subsequent year. All the other nine excluded schools were located in places that were geographically inaccessible, and financially too expensive to reach at the time of the survey. The within-school exclusion occurred chiefly because of the presence of some continuing education and private candidates in the mainstream schools.

4.4.6 Sample Size

Four main factors were considered in deciding the sample size for this study: the intended use of the findings, the characteristics of the defined target population, the budgetary limitations, and the administrative logistics.

It has been described in Chapter 1 that the main aim of the study is to design a national educational assessment model that will generate reliable information for use in making research-based decisions by stakeholders in the Bhutanese education system, especially policy makers. Given the importance of the information processed from the survey, the size of the sample should be such that it provided accurate and reliable information about the population parameters of interest. In other words, the population parameters obtained from the research should have the minimum possible error bound. Usually the error bound is set at two standard deviations of the estimator of interest (Scheaffer, et al., 2006), as depicted in *Equation* 4.1.

$$2\sqrt{V(\bar{y})} = B \qquad (4.1)$$

In *Equation* 4.1, V is the variance of estimator, $\bar{y}$, and B is the bound on the error of estimation. The variance, V, is related to population size, N, sample size, n, and variance, σ^2, respectively, as follows:

$$V(\bar{y}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) \qquad (4.2)$$

Substituting V in *Equation* 4.1 from *Equation* 4.2 leads to *Equation* 4.3:

$$2\sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)} = B \qquad (4.3)$$

Finally, $n = \dfrac{4N\sigma^2}{(N-1)B^2 + 4\sigma^2}$ (4.4)

Equation 4.4 shows that the sample size, n, is the function of a desirable error bound, B, and population variance, σ^2, respectively. *Equation* 4.3 shows that the error bound is inversely proportional to the square root of the sample size. Therefore, for a small error bound, a large sample size is required. However, large sample sizes often result in low sampling efficiency due to increased administrative costs and time. In view of this relationship between error bound and sample size, sampling experts choose a sampling design that would yield high sampling accuracy and efficiency equal to the sampling accuracy and efficiency that would have been obtained if a simple random sampling design was used (Kish, 1995; Ross, 2005; Scheaffer, et al., 2006). Ross (K. N. Ross, 1992; Ross, 2005) plausibly notes that the sampling accuracy of a large-scale educational survey is usually measured in terms of a simple random sample of 400 students, which is known as the effective sample size. This sampling accuracy would yield a 95% confidence interval for an estimate of a student-level mean of ±10 score points or a standard error estimate of approximately 5 points (Mullis, et al., 2008; K. N. Ross, 1992). Therefore, the sampling accuracy of the stratified two-stage cluster sampling design used in this thesis was required to have the same sampling accuracy of a simple random sample of 400 students. Similar designs were also used for TIMSS (IEA, 2004) and PISA (OECD, 2009b)

When a cluster of students are sampled from the same school, it is very likely that the students are of similar characteristics because they may live in the same geographical areas, influence each other through peer interactions, and take lessons from the same teachers. Such commonality reduces the possibility of random student enrolments in schools, thereby making them homogeneous in terms of certain characteristics. The presence of homogeneity within a cluster increases the variance of the sample. The measure of homogeneity within a cluster is *roh*, the coefficient of intra-class correlation (Kish, 1995; K. N. Ross, 1992; Ross, 2005). Assuming that the clusters are of equal size, Kish (1995) noted that the design effect, that is, the ratio of the variance of a sample to the variance of a simple random sample of the same number of elements, expresses the effect due to the homogeneity of the sample clusters. Kish (1995, p. 259) expressed this relationship as follows:

$$Deff = \frac{V(\overline{X}_c)}{V(\overline{X}_{srs})} = 1 + (b-1)roh \qquad (4.5)$$

In *Equation* 4.5 *Deff* is the design effect, $V(\overline{X}_c)$ is the variance of a complex sample (e.g., two-two stage sample design), $V(\overline{X}_{srs})$ is the variance of a simple random sample, *b* is the size of the selected clusters, *roh* is the coefficient of intra-class correlation, respectively. Ross (2005, p. 19) derived the relationship between the actual sample size, n_c ,and the effective sample size, n^*, from *Equation* 4.5 as follows:

$$n_c = n^*[1+(b-1)roh] \qquad (4.6)$$

Equation 4.6 enables a sampler to determine the actual sample size of a complex sample design in terms of the sampling accuracy of the estimates equivalent to that of a simple random sample of 400 students, effective sample size, when the value of *roh* is available. However, values of *roh* are often difficult to obtain, either because of the lack of past research in a similar area with a similar population, or due to high costs and time involved in piloting the research instruments with sample size sufficient to get the values of *roh*. In such situations, the value of *roh* is obtained from other similar research works (Kish, 1995) . The values of *roh* used in TIMSS (IEA, 2008) and PISA (OECD, 2009b) were referenced to calculate the sample size for the stratified two-stage cluster sampling design used in this thesis. While TIMSS (IEA, 2008) used *roh* of 0.3, PISA (OECD, 2009b, p. 204) presented 0.36, 0.34, and 0.36 as the median values of *roh* for the mathematics domain for the past three rounds of assessments . Since the values of *roh* used in TIMSS and PISA are very close to each other, 0.36 was used as the value of *roh* in *Equation* 4.6 to calculate the sample size for the stratified two-stage cluster sampling design. Based on the effective sample size of 400 students and the value of *roh*=0.36, a sample design table (Table 4.2) was constructed to determine the number of students and schools required for the stratified two-stage cluster sampling design.

Table 4. 2. Sample Design Table

Students per School	*roh*=0.1		*roh*=0.2		*roh*=0.3		*roh*=0.36	
	No. of Schools	Total Sample Size	No. of Schools	Total Sample Size	No. of Schools	Total Sample Size	No. of Schools	Total Sample Size
1	400.0	400	400.0	400	400.0	400	400.0	400
2	220.0	440	240.0	480	260.0	520	272.0	544
3	160.0	480	186.7	560	213.3	640	229.3	688
4	130.0	520	160.0	640	190.0	760	208.0	832
5	112.0	560	144.0	720	176.0	880	195.2	976
6	100.0	600	133.3	800	166.7	1000	186.7	1120
7	91.4	640	125.7	880	160.0	1120	180.6	1264
8	85.0	680	120.0	960	155.0	1240	176.0	1408

9	80.0	720	115.6	1040	151.1	1360	172.4	1552
10	76.0	760	112.0	1120	148.0	1480	169.6	1696
11	72.7	800	109.1	1200	145.5	1600	167.3	1840

Table 4.2 shows that as the number of students selected per school increases, the corresponding number of schools required to get the effective sample size decreases in a diminishing pattern. For the same number of students selected per school, the number of schools increases as *roh* increases.

Given that Bhutan had only 71 schools offering Grade 10-level education in 2009, and the likelihood that the *roh* is 0.36, Table 4.2 cannot be used in calculating the effective sample size for Bhutan. For example, as per *Equation* 4.6, that underpins Table 4.2, 145 schools are required to get the sample size that would yield the sampling accuracy equivalent to that of the effective sample size even when 200 students are selected from each one of the 71 schools, which is not possible in Bhutan. However, Table 4.2 provided helpful information in deciding the sample size for this thesis.

It is apparent that the number of schools offering Grade 10-level education in Bhutan in 2009 was insufficient for determining an effective sample size in accordance with the procedures entailed in Table 4.2. However, Kish (1995, p. 167) defined the population variance of the mean for a multi-stage cluster sample design as the sum of the variances at different stages of clustering. Applying Kish's relationship to a two-stage cluster sample design results in *Equation* 4.7.

$$Var(\bar{y}) = \left(1 - \frac{a}{A}\right)\frac{\sigma_a^2}{a} + \left(1 - \frac{b}{B}\right)\frac{\sigma_b^2}{ab} \qquad (4.7)$$

In *Equation* 4.7 $\bar{y}$ is the population mean, A is the total number of clusters, a is the number of clusters selected from A clusters, B is the total number of elements in cluster A, b is the number of elements selected from B, σ_a^2 is the variance of a clusters, and σ_b^2 is the variance of b elements. As per *Equation* 4.7, selecting all the schools (i.e., cluster A) would partial out σ_a^2 from the equation, leaving σ_b^2 as the only source of variance in the population mean. Since the standard error in the population mean is proportional to the variance, reducing the variance reduces the amount of standard error in the population mean. Further, *Equation* 4.8 expresses *rho* in terms of variances as follows (Kish, 1995, p. 170):

$$rho = \frac{\sigma_a^2 - \sigma_b^2 /(B-1)}{\sigma^2}, \tag{4.8}$$

where σ^2 is the variance on the population mean.

Equation 4.8 shows that the decrease in σ_a^2 leads to the decrease in *rho*, thereby minimizing the sampling error due to design effect. Based on the above observations, all the 60 schools included in the defined target population were sampled at the first stage to eliminate the between-school variance, σ_a^2, from the population mean. The idea of selecting all schools was also followed by TIMSS (IEA, 2008) and PISA (OECD, 2009b) in the countries that had fewer schools than the number of schools required to achieve the sampling accuracy equivalent to that of the effective sample size.

4.4.7 The First Stage Sampling Unit: Schools

All 60 schools in the defined target population were selected at the first stage. Because all 60 schools in the defined target population were sampled, the number of schools in the defined target population was equal to the sum of the school weights in the sample—a result that is usually achieved by following systematic random sampling procedures. For the same reason, stratification variables were not used in selecting schools at the first stage of sampling. However, the schools were weighted with probability proportional to their sizes to ensure that larger schools had a higher probability of selection than smaller schools and that students in larger schools had a smaller probability of being selected than students in smaller schools. This results in a uniform final student weighting that is essential for reducing sampling variance. Details of the various weights and the procedures used in calculating them are available in Appendix 1.

4.4.8 The Second Stage Sampling Units: Students

Selection of students from each one of the 60 sampled schools was influenced by two main factors. First, the research outcomes required that statistics be generated to compare student outcomes between schools. Second, the procedural validity in the data collection required that the number of students selected per school be small enough to facilitate constant vigilance and supervision by test administrators during the test session. Guided by these requirements, 25 students were randomly sampled from each of the 60 schools.

The group size of 25 students per school was deemed sufficient for comparing schools based on a similar size being used in other similar large-scale assessments (IEA, 2008; OECD, 2009b). In addition, the group size of 25 students per school enabled the test administrators to administer the test and the student questionnaire in a fair and just manner that ensured the procedural validity in data collection. For instance, test malpractices were neither detected when evaluating the test answer scripts nor were such incidents observed by the test administrators when administering the test.

A series of sequential steps were followed in sampling schools and students. From the sampling frame described earlier, a student-sampling frame of the Grade 10 students enrolled in 60 schools was constructed by listing all Grade 10 students enrolled in the 60 schools. The student-sampling frame was developed as follows. First, a four-digit school identification number was assigned to each of the 60 schools. The first two digits of the four-digit school identification number were to identify the school district and the last two digits were to identify the school. Second, a three-digit student identification number was assigned to each student enrolled in each of the 60 schools. Consequently, each student was assigned with a unique seven-digit identification number, with the first four digits providing information about his/her school and the last three digits providing his/her unique identification number. Therefore, the student-sampling frame provided access to the following information: the school district name, the school name, the student name, and the total number of students in a school. The seven-digit student identification numbers in the student-sampling frame were sequenced in ascending order.

To select 25 students from each one of the 60 schools, the last three digits of the seven-digit numbers were used. A group of 25 students from each of the 60 schools were sampled with equal probability of selection by using simple random sampling technique (Cochran, 1977; Kish, 1995; Scheaffer, et al., 2006). Students were weighted to ensure that they were selected in proportion to the size of their schools. The final student weights were calculated as the product of the school weights and the student weights. As shown in Appendix 1, the final student weights do not vary—a condition desirable for decreasing sampling variability (OECD, 2005d).

4.4.9 Replacement Students

In an ideal situation, it is natural to expect a 100% response rate, but practically it is difficult to achieve such a response rate. Therefore, it was anticipated that some of the 25 students sampled within each of the 60 schools would not be able to participate in the survey due to some problems. Accordingly, each school was allowed to replace three students.

The test administrators were instructed to replace a maximum of three students in each one of the 60 schools where some of the 25 students were not able to participate in the survey. The test administrators were provided with the student-sampling frame and the student tracking form. While the student tracking form was used for recording the attendance of the 25 sampled students on the day of the survey, the student-sampling frame was used in selecting replacement students. The test administrators were instructed to select the student replacement from the student sampling frame in such a way that the position of the replacement student immediately preceded the position of the student who was not able to participate in the survey.

Allowing schools to replace three students meant that the response rate from each school could be 88% without the use of replacement students, and 100% with the use of replacement students. The minimum response rate of 88% without using the replacement students is important for comparing schools and conserving the desired sample size. A similar restriction is followed in PISA, with the minimum acceptable response rate fixed at 85% (OECD, 2009b, p. 66).

4.5 Sample Design for the Focus Group Interview

A decision as to who should be invited to participate in a focus group interview depends on its purpose (Krueger & Casey, 2009; Morgan & Scannell, 1998). Further, in selecting participants for a focus group interview, the primary emphasis is on noting the possibility of sampling bias, not on generalizability (Morgan, 1988). The following paragraphs describe how these two aspects of sampling for a focus group influenced the selection of participants for the focus group interview used in this study.

The main purposes of the focus group interview were to collect perspectives and experiences of relevant stakeholders related to data required at the context-level of the proposed national educational assessment model. It was emphasized in the proposed model that the context for a nation's education system depends on external achievement stimuli, effective learning environments, and mechanisms for evaluating national education policies. The school education system in Bhutan is fairly centralised because the Ministry of Education prescribes the syllabi for school curriculum, deploys and promotes teachers, and funds the schools.

Given the centralized nature of the schools in Bhutan, policies developed by the Ministry of Education play a direct role in shaping the context of the school education system. Therefore, some Ministry personnel who had been involved in framing national education policies were invited to participate in the focus group interview in line with the principles of purposeful sampling (Krueger & Casey, 2009; Morgan, 1988; Morgan & Scannell, 1998). Three invitees agreed to participate in the focus group interview, and they consisted of almost an entire section of the Ministry of Education. Therefore, the roles and responsibilities of the section and the individual participants were treated as confidential in order to protect their identities from being disclosed, directly or indirectly. However, it must be noted that the participants had a long history of dealing with the matters related to school education policies developed by the Ministry. Therefore, their perspectives and experiences were considered a representative summary of the context of the Bhutanese school education system.

Because only one three-member focus group was involved, between-group analyses from the focus group interview are not possible.

4.6 Research Instrumentation

The following data collection instruments were used for the work associated with this study: a mathematics test, a student questionnaire, a teacher questionnaire, a school questionnaire, and a focus group interview. The following sections describe the procedures used in designing and developing these instruments.

4.6.1 Mathematics Test: Design and Development

As noted in Chapter 1, the main purposes of the Mathematics test were to generate: (a) profiles of Grade 10 Bhutanese students' knowledge of the school mathematics curriculum; (b) information on the students' preparedness to meet the challenges of the future; (c) information on the students' knowledge and skills suitable for adapting to rapid societal and technological change; and (d) international benchmarks for the students' mathematical knowledge and skills. In addition, the Mathematics test has been identified as the criterion variable for assessing the various educational effectiveness factors related to student, teacher, and school characteristics. With these purposes as guidelines, the Mathematics test was developed as follows.

Downing (2006) has summarised different test development procedures into the following 12 steps: overall plan; content definition; test specification; item development; test design and assembly; test production; test administration; scoring test responses; passing scores; reporting test results; item banking; and test technical report. In line with these 12 steps, a two-hour Mathematics test for Grade 10 Bhutanese students was constructed by using all of the 42 PISA mathematics items that were released by the OECD (2009c) for public use. These were used by the OECD in either PISA 2000 or PISA 2003. The original identity numbers of the 42 items with their corresponding sources are listed in Appendix 2.

A typical table of test specification provides the following information: subject contents and competencies assessed by the test, item formats used in the test, and the number of items used in the test (Black, 1999). Table 4.3 shows the test specification designed for the Mathematics test with the 42 PISA mathematics items in terms of the PISA Mathematical Literacy Framework.

Table 4. 3. Specification of the Mathematics Test in terms of the PISA Mathematical Literacy Framework

Mathematics Domain	Competency Cluster												Total
	Reproduction				Connections				Reflection				
	MR	CR	OR	TO	MR	CR	OR	TO	MR	CR	OR	TO	
Space and Shape	9a	5a 22a	1a	**4**	13a		1b 3a 3b 4a 24a	**6**			3c	**1**	**11**
Quantity	23b	14a 16a 16b 23a		**5**		12a 19a 23c		**3**			16c	**1**	**9**
Change and Relationship	7b 7c	26a	2a 6a 6c	**6**	7a 7d	6b 15a	2b	**5**		15b	26b	**2**	**13**
Uncertainty	18a	8a 17a		**3**	17b		11a 21a 25a	**4**	20a		10a	**2**	**9**
Total	5	9	4	**18**	4	5	9	**18**	1	1	4	**6**	**42**

MR=Multiple-choice response; CR=Closed-constructed response; OR=Open-constructed response; TO= Total number of items in a competency cluster

Table 4.3 provides the following information: total items on each content category, total items in each competency cluster, item competency levels, types of item formats, and positions of items in the test. For instance, there are 11 items assessing Space and Shape, with four items requiring reproduction competency, six items requiring connections competency, and one item requiring reflection competency. Out of these 11 items, two are multiple-choice items bearing question numbers 9a and 13a, two are closed-constructed response items bearing question numbers 5a and 22a, and seven are open-ended constructed response items bearing question numbers 1a, 1b, 3a, 3b, 4a, 24a, and 3c. In total, there are 42 test items in the test.

As PISA has released only 42 items, all of them were used in developing the Mathematics test so that the test assessed a range of mathematical knowledge and skills entailed in the Grade 10 mathematics curriculum followed by Bhutanese schools.

4.6.1.1 Test Validity

The American Educational Research Association, (AERA), the American Psychological Association, (APA), and the National Council on Measurement in Education (NCME), (1999, p. 9) described validity as the "degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests". This definition emphasises validity as a unitary concept in terms of evidence, and extends over construct validity (Aiken & Growth-Marnat, 2006; Gronlund & Waugh, 2009; Linn, 2002), content validity (Aiken & Growth-Marnat, 2006; Gronlund & Waugh, 2009; Linn, 2002), and criterion validity (Aiken & Growth-Marnat, 2006; Gronlund & Waugh, 2009; Linn, 2002).

AERA, APA, and NCME (1999) identified the following sources of validity evidence: contents of the test; response processes/patterns observed in the test; internal structure of the test; relation of the test to external criterion; and the consequences of testing. Relating these sources to PISA, similar evidence was compellingly described in its assessment frameworks (OECD, 1999b, 2004b, 2006), reports (OECD, 2000, 2004a, 2007a), and technical reports (OECD, 2002, 2005d, 2009b). Where validity evidence has been found to be weak or wanting in PISA, scholarly debates on such issues have already been discussed in Chapter 2 of this thesis. As described in that chapter, a serious issue with PISA is related to its alignment with school curricula used by participating schools. This issue is pertinent to the Mathematics test used in this study as well, because the test was constructed by using the PISA mathematics items.

A test with good content validity elicits a range of responses that are representative of the entire domain of knowledge and skills that the test is designed to measure (Aiken & Growth-Marnat, 2006). To ensure that the Mathematics test measured the mathematical knowledge and skills that Grade 10 Bhutanese students were expected to learn, a content validation study was conducted by evaluating the alignment of the 42 PISA items with the Grade 10 mathematics curriculum standards. The PISA Mathematical Literacy Assessment Framework has the following four overarching content domains: (a) space and shape, (b) quantity, (c) change and relationship, and (d) uncertainty. These domains were compared with the Grade 10 mathematics curriculum standards. The comparison was expected to indicate similarities and differences between the PISA Mathematical Literacy Assessment Framework and the Grade 10 mathematics curriculum. Similarities would indicate that the PISA mathematics items could assess the mathematical knowledge and skills of Grade 10 Bhutanese students, while differences would indicate otherwise. The following three designs were used in the comparison: Bhutanese Grade 10 mathematics curriculum objectives by PISA Mathematical Literacy Assessment Framework objectives; Bhutanese Grade 10 mathematics curriculum strands by PISA Mathematical Literacy Assessment Framework domains; and the PISA mathematics test items by the Bhutanese Grade 10 mathematics curriculum strands and standards.

First, the comparison between the objectives of Bhutanese Grade 10 mathematics curriculum and of the PISA Mathematical Literacy Assessment Framework was made as follows. The PISA Mathematical Literacy Assessment Framework expects 15-year-olds to “analyse, reason, and communicate ideas effectively as they pose, solve, and interpret mathematical problems in a variety of situations” (OECD, 2004b, p. 24). Similarly, Bhutanese Grade 10 mathematics curriculum expects Bhutanese students to reason, communicate, and confidently use their mathematical knowledge and skills as they solve, describe, explore, and discover mathematical problems in various situations (CAPSD & BBE, 2007). The use of common key words such as reason, communicate, solve mathematical problems, and situations in both the PISA Mathematical Literacy Assessment Framework and the Grade 10 mathematics curriculum indicates a high degree of similarity in their objectives. PISA expects students to demonstrate their capacities to analyse, reason, and communicate mathematical ideas effectively through their performances on the PISA mathematics test items that model the various real-world mathematical problem situations that the students are likely to encounter after their school days. Similarly the Grade 10 mathematics curriculum expects Bhutanese students to be able to “relate mathematics to the situations relevant to their daily lives” (CAPSD & BBE, 2007, p. 63). PISA is designed to assess the cumulative yield of education systems at the level of compulsory schooling, focussing on the knowledge and skills that students need in their adult life (OECD, 2004b, 2006). Similarly, the Grade 10-level school education marks the end of the basic education in Bhutan (Policy and Planning Division, 2008). In other words, the Grade 10-level school education is expected to prepare students either for further studies or provide them with the knowledge and skills relevant to their adult life. The inference that can be drawn from the comparison is that the objectives of the Bhutanese Grade 10 mathematics curriculum and of the PISA Mathematical Literacy Assessment Framework are satisfactorily similar—both expect students to learn and demonstrate similar mathematical knowledge and skills.

Second, the strands of the Bhutanese Grade 10 mathematics curriculum and the domains of the PISA Mathematical Literacy Assessment Framework were compared, as shown in Table 4.4.

Table 4. 4. Comparison of the Bhutanese Grade 10 Mathematics Curriculum Strands and the PISA Mathematical Literacy Assessment Framework Domains

Bhutanese Grade 10 Mathematics Curriculum Strands	PISA Mathematical Literacy Assessment Framework Domains			
	Space and Shape	Change and Relationship	Quantity	Uncertainty
10A: Numbers				
10B: Operations				
10C: Pattern				
10D: Measurement				
10E: Geometry				
10F: Data Management & Probability				

Notes: 1. White colour= does not match; 2. Blue colour= does match only in terms of the 42 PISA mathematics items; 3. Green colour= does match only in terms of the similarities in mathematical knowledge and skills; and 4. Purple colour= does match both in terms of the 42 PISA mathematics items and the similarities in mathematical knowledge and skills

In Table 4.4, the shaded cells indicate the match between the strands of the Bhutanese Grade 10 mathematics curriculum and the domains of the PISA Mathematical Literacy Assessment Framework. First, the mapping is done based on the similarities in mathematical knowledge and skills emphasized in the Bhutanese Grade 10 mathematics curriculum strands and the PISA Mathematical Literacy domains. For instance, Numbers in the Bhutanese Grade 10 mathematics curriculum emphasises the following mathematical knowledge and skills: an understanding of number meanings, ordering and representing real numbers, and applying a variety of number theory concepts in solving problems. Quantity in the PISA Mathematical Literacy Assessment Framework emphasises the following skills and knowledge: an understanding of relative size, the recognition of numerical patterns, the use of numbers to represent quantities and quantifiable attributes of real-world objects, and estimation (OECD, 2004b). As mathematical knowledge and skills emphasized in Numbers and Quantity are similar, they are mapped together. Second, the Bhutanese Grade 10 mathematics curriculum strands and the PISA Mathematical Literacy domains are mapped by allocating the 42 PISA mathematics items to the Bhutanese Grade 10 mathematics curriculum strands based on how close the items measured the mathematical knowledge and skills embedded in the strands. The results from the comparison, as displayed in Table 4.4, lead to the inference that the strands of the Bhutanese Grade 10 mathematics curriculum matched with the domains of the PISA Mathematical Literacy Assessment Framework. Therefore, it was possible to develop a test specification for the Mathematics test in terms of the Grade 10 mathematics curriculum strands. Table 4.5 shows the test specification.

Table 4. 5. Test Specification of the Mathematics Test in Terms of the Bhutanese Grade 10 Mathematics Curriculum Strands

Mathematics Curriculum Strands	Competency Cluster												Total
	Reproduction				Connections				Reflection				
	MR	CR	OR	TO	MR	CR	OR	TO	MR	CR	OR	TO	
10A: Numbers			2a	**1**									**1**
10B: Operations		16a 26a		**2**			2b	**1**			26b	**1**	**4**
10C: Pattern	7c 7b	5a 14a 16b	6a 6c	**7**	7a 7d	6b	3a 3b 24a	**6**			3c 16c	**2**	**15**
10D: Measurement	23b	22a 23a	1a	**4**	13a	19a 23c	1b 4a	**4**					**9**
10E: Geometry	9a			**1**									**1**
10F: Data Management & Probability	18a	8a 17a		**3**	17b	12a 15a	11a 21a 25a	**6**	20a	15b	10a	**3**	**12**
Total	5	9	4	**18**	4	5	9	**18**	1	1	4	**6**	**42**

The similarities in the strands of the Bhutanese Grade 10 mathematics curriculum and the domains of the PISA Mathematical Literacy Assessment Framework indicate that the 42 PISA mathematics items may adequately assess the mathematical knowledge and skills of Grade 10 Bhutanese students. However, only similarities in objectives and domains are not sufficient to conclude that a set of test items designed for assessing the PISA Mathematical Literacy Assessment Framework domains can be used for testing the Bhutanese Grade 10 mathematics curriculum strands. It is imperative that the 42 PISA items aligned well with the Bhutanese Grade 10 mathematics curriculum standards.

Webb (1997, 1999, 2006) convincingly used the following four criteria to assess the alignment between test items and curriculum standards: categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance of representation. Webb (1999) defined each of these four criteria as follows:

> The criterion of categorical concurrence between standards and assessment is met if the same or consistent categories of content appear in both documents. …. Depth-of-knowledge consistency between standards and assessment indicates alignment, if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards. …. The range-of-knowledge criterion is used to judge whether a comparable span of knowledge expected of students by a standard is the same as, or corresponds to, the span of knowledge that students need in order to correctly answer the assessment items/activities. …. The balance-of-representation criterion is used to indicate the extent to which items are evenly distributed across objectives. (pp. 7-8)

These alignment criteria were applied in studying the alignment between the 42 PISA mathematics items and the Bhutanese Grade 10 mathematics curriculum standards. A one-day alignment study was conducted in Bhutan to study the alignment between the 42 PISA mathematics items and the Bhutanese Grade 10 mathematics curriculum standards. Five mathematics teachers, who had been teaching the Bhutanese Grade 10 mathematics curriculum in five schools in Bhutan for over a range of five to 10 years, were involved in the alignment study. Each one of the five teachers was provided with a copy of the Grade 10 mathematics curriculum standards a week before the alignment study, with the instruction to review them. The objective of the advance distribution of the curriculum standards was to enable the teachers to recapitulate their experience with the curriculum standards, and prepare them for the alignment study.

On the day of the alignment study, the teachers were presented with the depth-of-knowledge levels used in the 42 PISA mathematics items. PISA (OECD, 1999b, 2004b, 2006) used the following three broad mathematical competency clusters to evaluate the depth-of-knowledge levels: reproduction cluster, connections cluster, and reflection cluster. These clusters were described in Chapter 2.

Each one of the five teachers was instructed to study the competency clusters and add their own additional clusters or descriptions under any cluster allocation. This activity helped the teachers to understand the competency clusters, and find out if they agreed to the clusters. The teachers were then instructed to discuss, in a group of two or three, their understanding of the competency clusters. The discussion further clarified the meanings of the clusters and enabled teachers to reach consensual understanding of the clusters. It is worth noting here that none of the five teachers proposed any additional clusters or additional descriptions within any cluster, indicating the adequacy of the clusters in covering a range of mathematical knowledge and skills expected of Bhutanese Grade 10 mathematics students. The teachers were instructed to assign individually a depth-of-knowledge level for each objective for each standard. Where an objective contained two or more depth-of-knowledge levels, the teachers were instructed to assign the higher level because the lower level is a prerequisite for the higher level. Next, the teachers were instructed to discuss the depth-of-knowledge levels that they assigned for each objective for each standard, with the goal of reaching consensus. The depth-of-knowledge levels assigned for each objective for each standard by the teachers were coded on a coding matrix prepared prior to the the alignment study. Table 4.6 shows a portion of the Grade 10 mathematics curriculum standards assigned with the depth-of-knowledge levels. The complete table is available in Appendix 3.

Table 4. 6. The Bhutanese Grade 10 Mathematics Curriculum Standards with the Depth-of-Knowledge Levels

Standards			Depth-of-Knowledge Level		
Strands	Goals	Objectives	R-1	C	R-2
10-A					
	3				
		10-A1	1	0	0
		10-A2	0	1	0
		10-A3	1	0	0
		10-A4	0	1	0
		10-A5	0	1	0
		10-A6	0	1	0

In Table 4.6, R-1 represents the reproduction cluster, C represents the connections cluster, and R-2 represents the reflection cluster. The Grade 10 mathematics curriculum has a hierarchical structure, with objectives forming the first level, goals the second level, and standards the third level. Table 4.6 is read as follows: standard A of Grade 10 (i.e., coded as 10-A) mathematics curriculum has three goals (i.e., quantified as 3) and six objectives (i.e., coded as 10-A1, 10-A2, …, 10-A6), with objective10-A1 only mapped to the R-1 level (i.e., indicated by the digit 1), and not mapped to any other levels (i.e., indicated by the digit 0). The Grade 10 mathematics curriculum has a total of six standards, 29 goals, and 61 objectives. The results of the activity of assigning a depth-of-knowledge level for each objective for each standard were two-fold (Webb, 1999). First, the activity made the teachers more familiar with what students were expected to know and do for each one of the objectives within a standard. Second, the assigned levels could be used as benchmarks for comparing the depth-of-knowledge level of each one of the 42 PISA mathematics items.

The activity of assigning a depth-of-knowledge level for each objective for each standard was followed by the activity of assigning a depth-of-knowledge level for each of the 42 PISA mathematics items. The teachers were asked to assign individually a depth-of-knowledge level for each of the 42 PISA mathematics items, with the instruction on how to use the coding matrix provided to them for recording their observations. Table 4.7 shows the coding matrix used for the activity. A sample of the coding matrix is presented in Appendix 4.

Table 4. 7. Coding Matrix for Assigning Depth-of-Knowledge Level to Test Items

Item Number	1a	1b	2a	2b	3a	3b	3c	4a	.	.	.	25a	26a	26b
Depth-of-Knowledge			R-1			C	R-2							

In Table 4.7, the alphanumeric numbers in the first row are the item numbers of the 42 PISA items. Table 4.7 should be read as follows: item 2a is assigned R-1, item 3b is assigned C, and item 3c is assigned R-2. Where an item involved more than one depth-of-knowledge level, the teachers were instructed to assign the higher level for the item. After all the teachers had completed the activity, they were asked to discuss the depth-of-knowledge levels assigned for each of the 42 PISA mathematics items, with the aim of reaching consensus. The consensus among the teachers is important because an item can have only one of the depth-of-knowledge levels assigned to it. Therefore, the teachers should agree on the most appropriate depth-of-knowledge level assigned to an item. After the teachers reached consensus, they were presented with the depth-of-knowledge levels assigned for each of the 42 PISA mathematics items. The teachers were asked to compare their consensual list with the PISA list, and evaluate the depth-of-knowledge levels assigned for each of the 42PISA mathematics items in the two lists. The comparison revealed that the two lists had the same depth-of-knowledge levels assigned for each item of the 42 PISA mathematics items, corroborating teachers' competency at evaluating the depth-of-knowledge levels inherent in each of the 42 PISA mathematics items. The activity served two purposes. First, the activity provided information about the teachers' competency at assigning the depth-of-knowledge levels for each of the 42 PISA mathematics items. Second, the activity ascertained that the teachers understood the depth-of-knowledge levels in their true sense, which essentially validates the depth-of-knowledge levels they assigned each objective for each standard of the Bhutanese Grade 10 mathematics curriculum.

The teachers were instructed to match the depth-of-knowledge levels of the 42 PISA mathematics items with the objectives of the Bhutanese Grade 10 mathematics curriculum standards, based on the condition that a student's response to the questions provided information about what the student knew or could do with respect to an objective (Webb, 1999). The teachers wrote each item's depth-of-knowledge level in each row of an objective corresponding to the item's column. Each objective that was matched to a depth-of-knowledge level of an item was called a hit. Multiple hits were allowed, and no limit on the number of hits for an item was set. This meant that an item could be matched to more than one objective. However, after discussing among themselves on their individual work, the teachers were able to reduce their differences in the number of hits for an item to a noticeable extent. Table 4.8 shows a portion of the coding matrix used by one of the teachers to match the depth-of-knowledge levels of the 42 PISA items to the objectives (complete table is available in Appendix 5).

Table 4. 8. Depth-of-Knowledge Level of Items Matched to the Bhutanese Grade 10 Mathematics Curriculum Objectives

Standards			Competency			Items										
						1a	1b	2a	2b	3a	3b	3c	4a	8a	9a	10a
S	G	O	R-1	C	R-2											
10-A																
	3															
		10-A1	1	0	0											
		10-A2	0	1	0											
		10-A3	1	0	0						C					
		10-A4	0	1	0			R-1								
		10-A5	0	1	0											
		10-A6	0	1	0											

In Table 4.8, objective 10-A4 is matched to item 2a which has R-1 depth-of-knowledge level. Similarly, objective 10-A3 is matched to item 3b which has C depth-of-knowledge level. These objectives have one hit each.

Table 4.8 was used for computing statistics for each standard of the Bhutanese Grade 10 mathematics curriculum on the four alignment criteria for content validation. In accordance with Webb's (1999) method, the number of hits was used to compute mean, frequency, and percentage to evaluate the alignment of test items and curriculum standards against the four alignment criteria. All of the statistics were computed for each teacher, and the mean of all the teachers was computed for each alignment criterion for use in evaluating the alignment of the 42 PISA mathematics items with the Bhutanese Grade 10 mathematics curriculum standards. The complete computations for each of the four alignment criteria are presented in Appendix 6. Table 4.9 presents the summary result of the alignment of the 42 PISA mathematics items and the Bhutanese Grade 10 mathematics curriculum standards.

Table 4. 9. Alignment of 42 PISA Mathematics Items and the Bhutanese Grade 10 Mathematics Curriculum Standards

Sl. No.	Standards	Categorical Concurrence	Depth-of-Knowledge Consistency	Range-of-Knowledge	Balance-of-Representation
1	10-Strand A-Numbers	No	Yes	No	No
2	10-Strand B-Operations	Yes	Yes	Yes	No
3	10-Strand C-Pattern	Yes	Weak	No	Yes
4	10 -Strand D-Measurement	Yes	Yes	No	Yes
5	10-Strand E-Geometry	No	Weak	Yes	Weak
6	10-Strand F-Data Management and Probability	Yes	Yes	Yes	Yes

In Table 4.9, *Yes* indicates alignment, *No* indicates non-alignment, and *weak* indicates a marginal alignment. Webb (1999) applied different conditions to each of the four criteria to judge the alignment of curriculum standards and assessments as described in the following lines. First, there should be at least six test items measuring content from a standard for there to be a categorical concurrence between the standard and the test. Second, for the depth-of-knowledge consistency between the standard and the test to exist, at least 50% of the test items corresponding to an objective had to be at or above the level of knowledge of the objective. Third, for the range of knowledge criterion of alignment to be acceptable, at least 50% of the objectives for a standard had to have at least one related test item. Finally, for the balance of representation criterion to be acceptable, an index of 0.70 or higher, computed based on the difference in the proportion of objectives and the proportion of hits assigned to the objective, is required. Depending on the percentage of alignment on an alignment criterion, Webb (1999) graded the overall alignment of a test and curriculum standards into fully aligned (100%), highly aligned (70% to 99%), partially aligned (50% to 69%) , and poorly aligned (less than 50%). These criteria were used in evaluating the alignment of the Mathematics test and the Bhutanese Grade 10 mathematics curriculum standards.

Overall, Table 4.9 shows that the Mathematics test and Bhutanese Grade 10 mathematics curriculum standards were partially aligned, indicating the validity of the Mathematics test in assessing Bhutanese Grade 10 students' mathematical knowledge and skills of their school mathematics curriculum. Specifically, 67% of alignment on the categorical concurrence, 67% of alignment on the depth-of-knowledge levels, 50% of alignment on the range of knowledge, and 50% of alignment on the balance of representation was achieved.

In summary, Bhutanese Grade 10 mathematics curriculum standards and the Mathematics test attained a sufficient level of alignment on all the four alignment criteria, namely, categorical concurrence, depth-of-knowledge consistency, range-of- knowledge correspondence, and balance of representation. Therefore, the inference is that the test contained sufficient items to assess the mathematical knowledge and skills of Grade 10 Bhutanese students in line with what they were expected to know and do according to the Grade 10 mathematics curriculum standards.

4.6.1.2 Field Administration: Trial

A two-hour Mathematics test, developed by using the 42 mathematics items from PISA, was trialled in a school in Bhutan with the view to improving the test before administering its final version. Thirty six students participated in the trial test.

The analysis of the trial test data focussed on the following characteristics of the Mathematics test: suitability of the writing time, unwanted ambiguity in the test item wordings, and adequacy of the answer space. In addition, students were invited to make post-test comments, particularly their feelings about the test. A complete protocol of the psychometric analysis of the test could not be applied in analysing the data from the trial test for want of an adequate sample size. For instance, a minimum sample size of five test candidates is recommended for every parameter of a test item such as difficulty index and discrimination index (Crocker & Algina, 2008; Nunnally, 1978). However, as the test items were adapted from PISA, information about 'item p-values' and 'item difficulty scores' was available (OECD, 2009c). In addition, the fact that the items were used in PISA is indicative of the items having sound psychometric properties.

First, the item p-values of the 42 items were examined. Table 4.10 presents the item p-values as reported by the OECD (2009c). An item p-value indicates the percentage of candidates who answered an item correctly. The item p-values range from zero to one, with the p-values less than 0.20 and greater than 0.95 indicating problems in the test items (Crocker & Algina, 2008). Table 4.10 also presents the item difficulty scores with a mean of 500 and a standard deviation of 100. An item difficulty score corresponds to the maximum score points on the PISA mathematics scale when the item is correctly answered.

Table 4. 10. Item p-values and Item Difficulties Reported by the OECD for the 42 PISA Mathematics Items

Item Number	Item p-value	Item difficulty	Item Number	Item p-value	Item difficulty
1 (a)	0.61	492	13 (a)	0.20	710
1 (b)	0.55	524	14 (a)	0.66	484
2 (a)	0.34	611	15 (a)	0.54	533
2 (b)	0.19	708	15 (b)	0.29	636
3 (a)	0.49	548	16 (a)	0.80	636
3 (b)	0.25	655	16 (b)	0.74	439
3 (c)	0.13	723	16 (c)	0.40	586
4 (a)	0.19	712	17 (a)	0.79	427
5 (a)	0.56	516	17 (b)	0.48	565
6 (a)	0.61	506	18 (a)	0.50	549
6 (b)	0.46	559	19 (a)	0.61	499
6 (c)	0.69	529	20 (a)	0.46	557
7 (a)	0.67	492	21 (a)	0.32	620

7 (b)	0.83	403	22 (a)	0.78	421
7 (c)	0.83	413	23 (a)	0.72	620
7 (d)	0.28	655	23 (b)	0.46	570
8 (a)	0.47	556	23 (c)	0.50	554
9 (a)	0.58	537	24 (a)	0.63	503
10 (a)	0.52	551	25 (a)	0.36	615
11 (a)	0.26	710	26 (a)	0.73	447
12 (a)	0.49	559	26 (b)	0.25	657

A histogram of the item p-values shown in Table 4.10 is presented in Figure 4.1. Figure 4.1 indicates that the item p-values are approximately normal distributed, showing that the Mathematics test was able to motivate and assess students of diverse abilities (Nunnally, 1978).

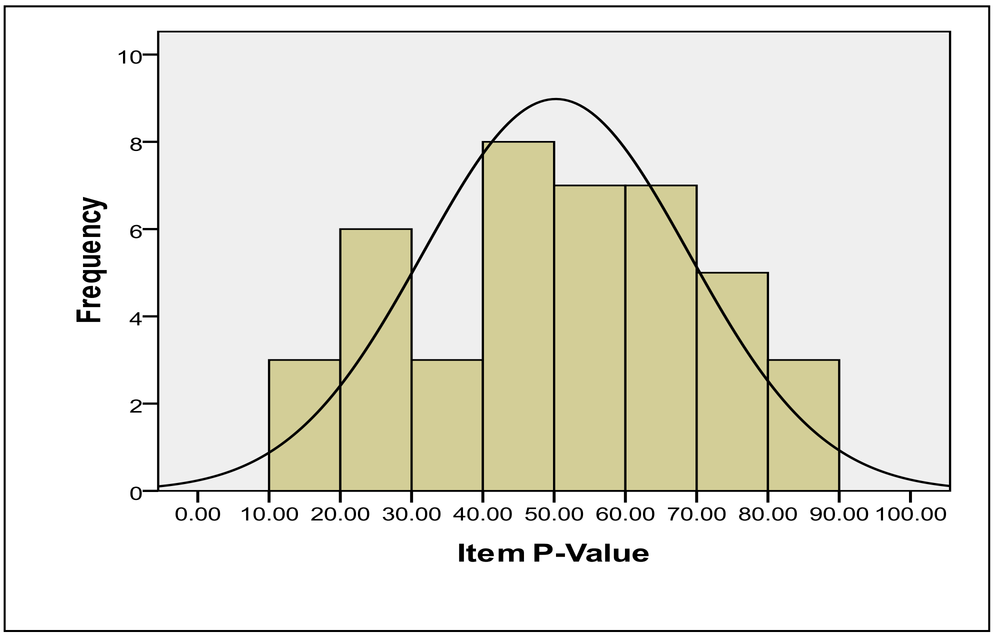


Figure 4. 1. Histogram of Item p-values

Second, the appropriateness of the test-writing time of two-hour duration was analysed by looking at the students' response patterns. Since the test did not contain choice within questions or alternate questions, it was assumed that the students would attempt to respond to as many questions as the writing time allowed them. Therefore, many answer papers with many incorrect or missing responses at the end of the test would suggest the inadequacy of the test-writing time. The answer paper did not show any discernible response patterns attributable to this assumption, proving that the test-writing time was in proportion to the mathematical knowledge and skills demanded by the test. In addition, to ensure that the test-writing time was not in excess of the mathematical knowledge and skills demanded by the test, students were encouraged to leave the test room immediately after they finished writing the test. It was observed that no students left the test room before the last five minutes of the test, confirming that the test-writing time was proportional to the mathematical knowledge and skills inherent in the test. The two-hour writing time was also in close proximity to the average of two-minute writing time for a test item followed in PISA (OECD, 2005d). The inference from these observations is that the test-writing time matched with the difficulty level of the test.

Third, the possibility of ambiguities in the item wordings of the test was analysed by studying the clarifications sought by the students during the test session. A few students sought clarification on some items. It was observed that some diagrams and graphs did not carry appropriate labels. Accordingly, this was corrected in the final version of the test.

Fourth, the adequacy of the writing space was analysed by studying the manner in which the students used the answer space provided in the test paper. No students had used a separate answer sheet for any question, and there was no spill over writing or writing in reduced font size in any answer space. This indicated the adequacy of the answer space provided for each item.

In summary, the trial test data showed that the Mathematics test of 42 items was suitable for use in testing Grade 10 Bhutanese students. The histogram of the item p-values also indicated that the test had acceptable items with an appropriate range of item difficulties to cater for diverse student abilities.

4.6.1.3 Field Administration: Final

The final administration of the Mathematics test consisted of informing the schools, training the administrators, and receiving the test materials from the test administrators. The edited version of the Mathematics test was administered to 1,500 students in 60 schools across Bhutan. Prior consent was obtained from the schools, and test schedules were provided to the schools before administering the test. The test was administered by a group of test administrators who were provided with a brief training on administering the test. The test administrators’ training focussed on the following areas: filling out the student tracer forms, setting up the test rooms, unpacking the test papers, administering the test, invigilating the test, packing the answer papers, and delivering the test documents to the researcher. The training was aimed at standardizing the test administration procedures so that the influence of different test administrators or test environments on the test candidates was avoided or reduced (Evers, 2001). In line with the training, the test administrators successfully administered the test in all 60 schools.

4.6.1.4 Test Scoring and Data Screening

Responses to the test were scored and the responses to the questionnaire items were coded before entering into computers. The test papers were evaluated and scored by the researcher in line with the test scoring guide (see Appendix 7) adapted from the one used for PISA (OECD, 2009c). This ensured consistency in the scores across the papers, and avoided the need for marker training and inter-rater agreement statistics. Responses to the questionnaires were similarly coded. The test scores and the responses to the questionnaires were entered into computers by a four-member data entry team. The team was trained on data entry protocols. During the training, each team member entered data from a sample of five test papers into a pre-designed data entry form. A similar procedure was followed with the questionnaires. The trial provided the team with hands-on experience and consolidated their prior verbal training. Data entry was done in the morning, followed by a data cleaning session in the afternoon. Data were cleaned by using descriptive statistics, such as the score frequencies, range, and mean. The errors detected during data screening were corrected by referring to the original response papers. This mode of data entry minimised the errors committed during entering data into computers, as evidenced by fewer errors after the first couple of days.

Finally, the complete data set was screened in the same manner, and all errors were rectified before accepting the complete data set from the data entry team. Throughout the data entry period, the researcher worked as the table leader for the data entry team. The complete data set was then made ready for further analyses.

4.6.1.5 Test Reliability

Classical test score theory assumes that each test taker has a true score that would be obtained if there were no errors in measurement (Aiken & Growth-Marnat, 2006; Kaplan & Saccuzzo, 2001; Nunnally, 1978). Errors in measurement can be systematic or random. Kaplan and Saccuzzo (2001) noted that systematic errors in measurement are less likely than random errors to misguide an investigator to make wrong inferences. Classical test score theory assumes that measurement errors are random, and are attributable to external factors such as the unconducive test environment, test fatigue, and unmotivated test participants. Therefore, a test score has two parts, namely, true score and random measurement errors. Test reliability indicates the proportion of variability attributable to the true score, that is, the percentage of the variance in the test score because of the existence of a true score. The higher the reliability, the lesser the variance in the test scores because of random errors.

The reliability estimate for the mathematics test used in this study was calculated by using Cronbach's coefficient alpha (Aiken & Growth-Marnat, 2006; Kaplan & Saccuzzo, 2001). Coefficient alpha was preferred over other methods of estimating reliability (e.g., test-retest, parallel forms, split-half, KR20, KR21) because it is "the most general method of finding estimates of reliability through internal consistency" (Kaplan & Saccuzzo, 2001, p. 113). In addition, the test administrators were adequately trained on the ways to prevent possible errors related to test administration and logistics from unduly affecting students when writing the test. Further, only one form of the test was used in all the schools. In this way, the errors associated with test administration and different test forms were adequately controlled. Therefore, the reliability estimates that can be obtained by other methods are less relevant to the test than the internal consistency. The coefficient alpha for the Mathematics test was 0.78, with the standard error of 0.01 score units. TiaPlus (Heuvelmans, 2002) was used in estimating the coefficient alpha. This reliability estimate is deemed adequate for the test. Evers (2001) noted that a reliability coefficient between 0.70 and 0.80 is sufficient for making decisions on students' learning, and a reliability coefficient greater than 0.70 is good for research at group level; indicating the adequacy of the reliability coefficient of 0.78 for the test. Similar values have been reported for PISA (OECD, 2009b) and TIMSS (IEA, 2008).

4.6.1.6 Test Item Calibration and Student Ability Estimation

The Mathematics test items were calibrated by using ConQuest—a computer software program capable of generating item parameters based on both classical test theory and item response theory. Among the different item response models that ConQuest is capable of fitting (Adams & Wilson, 1996; Wu, Adams, Wilson, & Haldane, 2007), the partial credit model (Masters, 1982; Masters & Wright, 1997) was used in calibrating the Mathematics test items because of its close fit with the items used in the test. To facilitate the interpretation of the item parameters, the partial credit model is briefly discussed.

According to the partial credit model, the probability of person *j* scoring *x* on item *i* is written as

$$P_{ijx} = \frac{\exp \sum_{k=0}^{x_i} (\theta_j - \delta_{ik})}{\sum_{h=0}^{m_i} \exp \sum_{k=0}^{h} (\theta_j - \delta_{ik})}, \text{ where } \sum_{k=0}^{x_i} (\theta_j - \delta_{ik}) \equiv 0 \qquad (4.10)$$

In *Equation* 4.10, P_{ijx} is the probability of person *j* scoring *x* on item *i*, θ_j is the ability of person *j* or person parameter, and δ_{ik} is the item step difficulty or item parameter. The item parameter, δ_{ik}, is the relative difficulty of each step of an m_i step item with $m_i + 1$ step difficulties. In other words, the item parameter corresponds to a position on the measurement variable at which a person *j* of ability θ_j is equally likely to respond in either category *x* or category *x*-1. The person ability parameter, θ_j, is the modelled location of person *j* on the ability continuum. The partial credit model automatically takes the form of the Rasch model (Rasch, 1980) when an item has only two response categories.

The Mathematics test items were calibrated by using the maximum likelihood estimation procedure available in ConQuest (Wu, et al., 2007). Adams and Wilson (1996) present an excellent discussion on the use of the maximum likelihood estimation procedure in ConQuest, and the same is not repeated here. The result of the item calibration is useful for deciding whether an item should be excluded from the final analyses, because an item with poor psychometric properties provides misleading information. Table 4.11 shows the items from the Mathematics test with poor psychometric properties.

Table 4. 11. Items with Poor Psychometric Properties

Item ID	Estimate	SE	Unweighted Fit			Weighted Fit		
			MNSQ	CI	T	MNSQ	CI	T
Q7d	1.391	0.103	1.87	(0.93, 1.07)	19.0	1.11	(0.83,1.17)	1.3
Q13a	2.871	0.058	1.21	(0.93, 1.07)	5.3	1.03	(0.71,1.29)	0.2
Q23a	-1.234	0.031	1.15	(0.93, 1.07)	4.0	1.11	(0.95, 0.05)	4.0

Estimate=item difficulty parameter; SE=standard error of the item difficulty parameter; MNSQ=the mean square fit statistics; CI=95% confidence interval for the MNSQ; T=the Student's t-statistic.

For an item to fit the item response model used in the calibration, the value of the MNSQ should lie within the range of its corresponding CI, which was not the case with the items in Table 4.11. Also, when the value of the MNSQ does not fall within its corresponding CI, the absolute value of the corresponding t-statistic is greater than 2, which is indicative of the misfit between an item and the model used (Wu, et al., 2007), and this was the case with the items in Table 4.11. Therefore, the items in Table 4.11 did not fit the model. These misfit items were further examined by using their classical test theory-based indices generated by ConQuest. Table 4.12 shows the item indices of the items in Table 4.11 based on classical test theory.

Table 4. 12. Item Indices based on Classical Test Theory for the Items shown in Table 4.11

Item ID	% correct	Point Bi-serial correlation
Q7d	6.89	-0.10
Q13a	2.81	0.05
Q23a	43.11	0.27

In Table 4.12, item Q7d is a dichotomous item whose point bi-serial correlation is the same as its item-rest correlation of -0.10. The negative item-rest correlation shows that the item functioned differently from the rest of the items in the test. On closer inspection of the Mathematics test, the item had some typographical errors in its response options. The item was excluded from the analysis. Item Q13a has a very low point bi-serial correlation, showing that the item did not discriminate well between the performances of the high and the low ability students (Crocker & Algina, 2008). Only about 3% of students had achieved the maximum score on this item. However, the item neither had any typographical error nor it was irrelevant to the Grade 10 Bhutanese mathematics curriculum. Therefore, item Q13a was included in the later analyses. The point-bi-serial correlation of item Q23a is acceptable, and therefore it was included in the later analyses.

Besides the sound psychometric properties of the test items, students' motivation during the test and sufficiency of the test writing-time also affect the test validity. The missing response data were analysed next to confirm or rule out the possibility of the influence of these factors on students' responses to the test items. The analysis of the missing values in the test data by using SPSS version 19 (SPSS Inc, 2009) MVA (Missing Value Analysis) option showed that the data were MCAR (missing completely at random) as indicated by Little's MCAR test (χ^2=39103.219, df=39039, p=0.409). The 'Missing Patterns' table generated during MVA did not show any discernible response patterns. For instance, a not-reached item was expected to have an item immediately preceding it and the remaining items following it left unanswered, which was not visible in the 'Missing Patterns' table. Two inferences were drawn from these observations. First, the students remained motivated when writing the test. Second, the missing values in these data represented incorrect responses. In the light of these conclusions, the misfit of the items 13a and 23a to the model were attributed to their difficulty; not to the lack of sound psychometric properties of these items. Therefore, these items were included in the item calibration as well as in generating student proficiency scores. Thus, overall the Mathematics test was reduced from 42 items to 41 items.

Item calibration was then followed by estimation of student ability scores. A commonly used method of estimating student ability scores on large-scale assessment surveys is the use of plausible values (Beaton & Gonzalez, 1995; Mislevy, Beaton, Kaplan, & Sheehan, 1992; OECD, 2005c, 2009b; Wu, 2005). Similarly, student ability scores on the Mathematics test were plausible values that were generated by ConQuest, with the mean of zero and the standard deviation of one. A set of five plausible values was generated for each student. The plausible values were then transformed to the PISA scale with the mean of 500 score points and the standard deviation of 100 units. The transformation helped in interpreting the Mathematics test scores in terms of the PISA Mathematics Proficiency Scale and facilitated the comparison of the Mathematics test scores with the PISA 2003 Mathematical Literacy test scores.

4.6.1.7 Linking the Mathematics Test to PISA 2003

The Mathematics test and the mathematical literacy section of the PISA 2003 were linked based on the approach of Kolen and Brennan (2004). The approach involves using the means and the standard deviations of the common items in the tests that are to be linked. The means and the standard deviations of the parameters of the common items in the Mathematics test and the mathematical literacy section of the PISA 2003 were used in deriving the scaling constants.

It is beyond the scope of the thesis to describe fully the procedures involved in linking different tests. However an overview of Kolen and Brennan's (2004) approach is necessary to set the task in context. Kolen and Brennan (2004) used the following transformation equations for scale *I* (test I) and scale *J (test J)*:

$$\theta_{Ji} = A\theta_{Ii} + B, \tag{4.11}$$

$$b_{Ji} = Ab_{Ij} + B, \tag{4.12}$$

In *Equations* 4.11 and 4.12, θ denotes the student ability and *b* denotes the item difficulty parameter, respectively. The constants *A* and *B* are the scaling constants. Kolen and Brennan (2004), also presented the following equations for deriving the scaling constants:

$$A = \frac{\sigma(b_J)}{\sigma(b_I)}, \tag{4.13}$$

$$B = \mu(b_J) - A\mu(b_I), \tag{4.14}$$

In *Equations* 4.13 and 4.14, σ and μ are the standard deviation and mean of the item difficulty parameters. Johnson and Owen (1998) also used similar equations for linking NAEP and TIMSS results. Using the means and the standard deviations of the 11 common items shown in Table 4.13, the scaling constants are derived as follows:

$$\alpha = \frac{\sigma_{P_{03}}}{\sigma_{M_T}} = 0.983 \tag{4.15}$$

$$\beta = \mu_{P_{03}} - \alpha(\mu_{M_T}) = -0.040 \tag{4.16}$$

In *Equations* 4.15 and 4.16, $\mu_{P_{03}}$, $\sigma_{P_{03}}$, μ_{M_T} and σ_{M_T} denote the means and standard deviations of difficulty estimates of the 11 common items for PISA 2003 and the Mathematics test respectively. Table 4.13 presents the item parameters of the common items used in the Mathematics test and the PISA 2003 Mathematical Literacy test along with the derived scaling constants.

Table 4. 13. Common Items and their Parameters as Calibrated in the Mathematics Test and in PISA

PISA 2003 Mathematics Test			Mathematics Test		Difference	Difference Squared
Item ID	Difficulty Estimate	Centred Difficulty Estimate (1)	Difficulty Estimate	Centred Difficulty Estimate(2)	(1-2)	
Q1(a)	-0.867	-0.125	-0.924	-0.21	-0.085	0.007225
Q1(b)	-0.453	0.289	-0.265	0.449	0.16	0.0256
Q7(a)	-0.861	-0.119	-0.706	0.008	0.127	0.016129
Q7(b)	-2.037	-1.295	-2.126	-1.412	-0.117	0.013689
Q8(a)	0.101	0.843	0.134	0.848	0.005	0.000025
Q14(a)	-0.824	-0.082	-0.908	-0.194	-0.112	0.012544
Q15(b)	1.119	1.861	1.248	1.962	0.101	0.010201
Q16(a)	-1.833	-1.091	-1.511	-0.797	0.294	0.086436
Q16(b)	-1.408	-0.666	-1.200	-0.486	0.18	0.0324
Q16(c)	0.474	1.216	0.292	1.006	-0.21	0.0441
Q17(a)	-1.567	-0.825	-1.886	-1.172	-0.347	0.120409
Mean	-0.742		-0.714		-0.00036	
SD	0.988		1.005		0.19203	

Also, shown in Table 4.13 is a Difference column that indicates the changes in the item difficulty parameters across the two tests. In principle, it is desirable to have a perfect match between the item parameters of the common items in the two tests. However, the item difficulty parameters of the common items differ between the tests which leads to linking error (OECD, 2005c). Therefore, the linking error is the standard error of the difference in the item difficulty parameters of the common items across tests, and it is computed by dividing the standard deviation of the difference by the square root of the number of common items (OECD, 2005c). The standard deviation of the difference in the item difficulty parameters of the 11 common items is equal to 0.192. Therefore, the linking error is 0.059 logit units. This linking error corresponds to 5.79 on the Mathematics test with the mean of 500 and the standard deviation of 100. The linking error leads to the overestimation of the mean test scores. Elaborate discussions on the properties of the linking error are presented in Kolen and Brennan (2004) and the OECD (2005c). In line with its properties, the linking error was used in computing the variances of the sample statistics.

Equation 4.11 makes it possible to link the PISA 2003 Mathematical Literacy test to the Mathematics test as follows:

$$P_{03} = \alpha(M_T) + \beta \tag{4.18}$$

In *Equation* 4.18, P_{03} denotes the predicted scores on PISA 2003 Mathematical Literacy test for Grade 10 Bhutanese students, M_T denotes the scores of Grade 10 Bhutanese students on the Mathematics test, and α and β are the scaling constants.

Substituting the values of the scaling constants in *Equation* 4.18 yields the following equation:

$$P_{03} = 0.983(M_T) - 0.040 \tag{4.19}$$

Equation 4.18 is in logit metric. Because the PISA 2003 Mathematical Literacy scores were standardised with a mean of 500 scores and a standard deviation of 100 units, Equation 4.18 was also standardised as shown in *Equation* 4.20.

$$P_{03-s} = (0.983(M_T) - 0.040) \times 100 + 500 \tag{4.20}$$

Equation 4.20 makes it possible to compare the Mathematics test scores with the PISA 2003 Mathematical Literacy test scores, and use the PISA Mathematics Proficiency Sale to interpret Grade 10 Bhutanese students' Mathematics test scores. The OECD (2005d) presents a thorough technical discussion on scale transformation between different cycles of PISA administrations which further guided the formulation of *Equation* 4.20.

4.6.1.8 Developing the Mathematics Proficiency Scale

Students' profiles of mathematical knowledge and skills were analysed and interpreted in terms of the proficiency levels assigned to their scores. Item response theory modelling of students' response data makes it possible to estimate student ability and item difficulty parameters separately, and position the two parameters on the same measurement scale (Embretson & Reise, 2000). This makes it possible to map item difficulty with student ability. By defining the mathematical knowledge and skills demanded by individual test items, it then becomes possible to interpret student ability in accordance with the mathematical knowledge and skills included in the test items. This provides an analytical insight into the profile of students' mathematical knowledge and skills. Such an insight has a huge potential to assist teachers to teach better, students to learn with understanding, and school heads to make informed decisions.

An example of an item map from PISA (OECD, 2005d, p. 257), with a slight modification, is shown in Table 4.14 for item 2a of the Mathematics test. The OECD (2005d, 2009b) presented detailed information on setting up performance levels for PISA test items, including the items used in its Mathematical Literacy test.

Table 4. 14. Item Map for Item 2a Adapted from the OECD

Item ID	Item Difficulty	Comments-item demands
2a	611	Interpret and link picture, text and algebra; algebraic substitution; solve basic equation; single step; correct manipulation of expressions containing symbols

As the Mathematics test was adapted from PISA (OECD, 2009c) and the two were linked, the PISA 2003 Mathematics Proficiency Levels were used for interpreting the Mathematics test scores. PISA has six proficiency levels, ranging from 1 to 6, with each level having performance descriptions of the types of mathematical knowledge and skills that students attaining a particular level are likely to demonstrate or possess. Each proficiency level is assigned with performance scores proportional to the complexity of the items. Students are identified with different proficiency levels based on their performance scores vis-à-vis the scores corresponding to each proficiency level. Table 4.15 displays the PISA Mathematics Proficiency Scale and Levels.

Table 4. 15. The PISA Mathematics Proficiency Scale and Levels (OECD, 2005d, pp. 260-261)

Level	Score Points on the PISA Scale	Summary Descriptions for Six Levels of Overall Mathematical Literacy
6	Above 669	At Level 6 students can conceptualise, generalise, and utilise information based on their investigations and modelling of complex problem situations. They can link different information sources and representations and flexibly translate among them. Students at this level are capable of advanced mathematical thinking and reasoning. These students can apply their insight and understandings along with a mastery of symbolic and formal mathematical operations and relationships to develop new approaches and strategies for attacking novel situations. Students at this level can formulate and precisely communicate their actions and reflections regarding their findings, interpretations, arguments, and the appropriateness of these to the original situations.
5	607 to 669	At Level 5 students can develop and work with models for complex situations, identifying constraints and specifying assumptions. They can select, compare, and evaluate appropriate problem-solving strategies for dealing with complex problems related to these models. Students at this level can work strategically using broad, well-developed thinking and reasoning skills, appropriate linked representations, symbolic and formal characterisations, and insight pertaining to these situations. They can reflect on their actions and formulate and communicate their interpretations and reasoning.
4	545 to 607	At Level 4 students can work effectively with explicit models for complex concrete situations that may involve constraints or call for making assumptions. They can select and integrate different representations, including symbolic, linking them directly to aspects of real-world situations. Students at this level can utilise well-developed skills and reason flexibly, with some insight, in these contexts. They can construct and communicate explanations and arguments based on their interpretations, arguments and actions.
3	482 to 545	At Level 3 students can execute clearly described procedures, including those that require sequential decisions. They can select and apply simple problem-solving strategies. Students at this level can interpret and use representations based on different information sources and reason directly from them. They can develop short communications reporting their interpretations, results and reasoning.
2	420 to 482	At Level 2 students can interpret and recognise situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formulae, procedures, or conventions. They are capable of

		direct reasoning and making literal interpretations of the results.
1	358 to 420	At Level 1 students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry our routine procedures according to direct instructions in explicit situations. They can perform actions that are obvious and follow immediately from the given stimuli.

As shown in Table 4.15, the levels correspond to a certain range of ability scores, and the levels are described in terms of the complexity of mathematical knowledge and skills required to attain each level, with the complexity increasing in proportion to the order of the levels.

4.6.2 Student, Teacher, and School Questionnaires

One of the purposes of this study was to study how student, teacher, and school characteristics relate to student achievement. As described in Chapter 3, different educational effectiveness factors related to student, teacher, and school have been assigned to their corresponding levels of the national educational assessment model. Data on these factors were collected by administering a questionnaire each to students, teachers, and school principals.

The questionnaires were developed by using scale items from various past studies. Three criteria were followed in adapting questionnaire items from the past studies. First, the prospective questionnaire item must measure the educational effectiveness factors identified in the national educational assessment model. Second, the prospective questionnaire item must be an item used in different countries, so that the results from the studies validate its parameter invariance across a range of cultural settings. Third, the prospective questionnaire item must be an item replicated in different studies, so that the results prove its reliability. Drawing on these criteria, items from NAEP, TIMSS, PISA, and the *Teaching and Learning International Survey* (TALIS) were found suitable for use in this study; consequently, they have been adapted in student, teacher, and school questionnaires.

Two types of indices were constructed from data collected with the student, teacher, and school questionnaires. First, simple indices were constructed through arithmetical transformation or recoding of two or more items when the items measured definitive attributes (e.g., age, gender). Second, scale indices were developed for the construct of interest that had an underlying quantitative continuum (e.g., motivation, self-concept) by using the Item Response Theory scaling methodology described in PISA (OECD, 2005d). Simple indices do not need rigorous validation because the items measure definitive attributes, but scale indices do need rigorous validation because the items measure a construct of interest.

As scale indices involve a construct of interest with an underlying quantitative continuum, they should undergo construct validation. The correspondence between a construct and its measures indicates construct validity (Bagozzi & Phillips, 1982). Construct validation is a multi-step process involving different criteria. The criteria frequently used to assess construct validity are: unidimensionality; within-method convergent validity; reliability; stability; across-method convergent validity and discriminant validity; and nomological validity (J. C. Anderson & Gerbing, 1988; Learly-Kelly & Vokurka, 1998; Steenkamp & van-Trijp, 1991; Zimmerman & Martinez-Pons, 1988). However, not all of these criteria can be applied to validating the constructs used in the questionnaires. For instance, the stability criterion requires at least two rounds of administration of the same measures of the constructs, and the across-method convergent validity criterion involves different methods of measuring the same constructs, which are both beyond the scope of this study. The remaining criteria of the construct validity were assessed by performing confirmatory factor analyses (CFA). The LISREL programme (Joreskog & Sorbom, 2003) was used to estimate the models for both Likert-type and dichotomous items.

The use of LISREL in construct validation is widely reported in literature (J. C. Anderson & Gerbing, 1988; Arthaud-Day, Rode, Mooney, & Near, 2005; Bagozzi & Phillips, 1982; Learly-Kelly & Vokurka, 1998; OECD, 2009b; Steenkamp & van-Trijp, 1991). Steenkamp and van-Trijp (1991) presented a compelling CFA model as follows:

$$x = \Lambda \xi + \delta \tag{4.21}$$

In *Equation* 4.21, x is the $q \times 1$ vector of the *n* sets of observed variables, ξ is $n \times 1$ vector of the constructs, Λ is the $q \times n$ matrix of regression coefficients relating the observed variables to the constructs, and δ is the $q \times 1$ vector of error terms of the variables. E*quation* 4.21 was used as the CFA model throughout this chapter, unless specified differently.

Generally, the model parameters were estimated by using the maximum likelihood estimation (MLE) procedure using covariance matrices. Because the MLE assumes multivariate normality, data were screened for skewness and kurtosis. A few of the observed variables deviated from normal distribution. In such cases data were then transformed with the “Normal Scores” facility in LISREL, and the robust maximum likelihood method of estimation was used, which is known to be less sensitive to distributional violations (Arthaud-Day, et al., 2005; Hoogland & Boomsma, 1998). For dichotomous observed variables, the weighted least squares (WLS) estimation method and polychoric correlations were used to estimate the model parameters (Joreskog & Sorbom, 2002). The fit of the model, described in *Equation* 4.21, to the empirical data was judged by using the Root-Mean Square Error of Approximation (RMSEA), the Root Mean Square Residual (RMR), the Comparative Fit Index (CFI), and the Non-normed Fit Index (NNFI). The reasons for choosing these fit indices over other fit indices are convincingly described by Kline (2005) and Byrne (1998); therefore, they are not recalled in this study.

The overall fitness of the model in *Equation* 4.20 to the data is indicative of the unidimensionality of the observed variables; the observed variables measure only the construct that they are supposed to measure (J. C. Anderson & Gerbing, 1988; Hattie, 1985; Steenkamp & van-Trijp, 1991). For the observed variables to pass the criterion of convergent validity, their estimated pattern coefficients or regression coefficients on their underlying construct should be significant (J. C. Anderson & Gerbing, 1988; Steenkamp & van-Trijp, 1991), and preferably the correlation between the observed variable and the construct should be more than 0.50 (Steenkamp & van-Trijp, 1991). *Equation* 4.20 can also be used in assessing discriminant validity for two constructs by constraining the estimated correlation parameter between the constructs to unity, followed by a chi-square difference test on the values obtained for the constrained and the unconstrained models (J. C. Anderson & Gerbing, 1988; Bagozzi, Yi, & Phillips, 1991). Bagozzi and Phillips (1982, p. 476) noted that a “significantly lower χ^2 value for the model in which the trait correlations are not constrained to unity would indicate that the traits are not perfectly correlated and that discriminant index is achieved”. However, as the chi-square test depends on sample size, a large sample size may reject even good fitting models (Fornell & Larcker, 1981; Schmitt & Stults, 1986). An alternative method of assessing the discriminant validity, which is followed in this study, requires that the average variance explained across the measures be more than the shared variance between their underlying latent constructs (Arthaud-Day, et al., 2005; Fornell & Larcker, 1981).

Composite reliability of the observed variables, that is, the reliability of their underlying construct can be calculated by using the parameter estimates generated with *Equation* 4.21 as follows (Diamantopoulos & Siguaw, 2000; Fornell & Larcker, 1981).

$$\rho_c = (\sum \lambda)^2 / [(\sum \lambda)^2 + \sum (\theta)] \qquad (4.22)$$

In *Equation* 4.22, ρ_c is the composite reliability, λ is the $p \times 1$ vector of standardised factor regression coefficients, and θ is the error variances of the observed variables. Since the “square root of the reliability is an upper bound for validity”(Scheerens, et al., 2003, p. 116), a higher value of reliability is indicative of construct validity, and was used in the analysis.

Next, the following sections present procedures used in developing student, teacher, and school questionnaires, and describe their field administrations. The sections also present the results of the CFA for the scale indices constructed from questionnaire data.

4.6.2.1 The Student Questionnaire

The student questionnaire was developed by adapting items from NAEP, TIMSS, and PISA. With their histories of repeated administration across different countries or states, and similarities between their purposes and the purposes of the proposed national educational assessment model, NAEP, TIMSS, and PISA questionnaire items satisfied the four criteria mentioned earlier. The complete details of the student questionnaire items and their sources are provided in Appendix 8. Table 4.16 shows the constructs and their corresponding items, with the numbers in the parentheses indicating the item numbers. These are not a one to one match for student variables allocated at the student-level of the proposed national educational assessment model, but they are all there in some form.

Table 4. 16. Constructs and Items used in the Student Questionnaire

Constructs	Questionnaire Items (Item Number)
Student background	Gender (1); Age (2); Grade repetition (22a-c); Language used at home (11a-d)
SES	Mother's occupation (3 & 4); Mother's highest level of schooling (5a-e); Mother's qualification (6a-c); Father's occupation (7 & 8); Father's highest level of schooling (9a-e); Father's qualification (10a-c); Family wealth (13a-e)
Educational resources	Home educational resources (12a-n); Books at home (14a-f)
Homework	Frequency of homework (15a-e); Homework time (16a-e); Study time (17a-f); Kinds of out-of-school-time lessons (18a-f)
Engagement	Self-efficacy (19a-h); Anxiety (20a-j); Learning situations (21a-j); Expected education level (23a-c); Attitudes towards school (24a-d); Reasons for choosing school (25a-f); Student and teacher relation (26a-e); Sense of belongingness (27a-f); Arriving late for the school (28a-d); Teacher support and discipline climate (29a-k); Motivations (30a-h); Learning strategies (31a-n)
ICT	Availability of computers (32a-c); Experience with computers (33 & 34a-d); Places for using computers (35a-c); Use of computers (36a-l); Confidence in ICT use (37a-w); Attitudes towards ICT (38a-d); Teaching Computer (39a-e); Teaching Internet (40a-f)

4.6.2.1.1 Field Administration

The 40-item student questionnaire was administered to Grade 10 Bhutanese students who sat the Mathematics test. Unlike the Mathematics test, the student questionnaire was not trialled in Bhutanese schools. However, additional measures were incorporated in the student questionnaire to compensate for the information that could have been obtained from trialling. The student questionnaire was administered by the same test administrators who administered the Mathematics test. The minimum time for completing the questionnaire was 35 minutes. However, the test administrators were instructed to give extra time to students who required more time to complete the questionnaire. The test administrators were also instructed to explain the meaning of words or parts of the questionnaire items as and when students sought such help. Special instructions were provided to students on the questionnaire that invited them to seek clarifications on words or questionnaire items. This was done to ensure that students understood the questionnaire items that had not been modified through trial testing the questionnaire. It may be worth recalling that only 25 students from each one of the 60 schools were selected to participate in the study, which enabled test administrators to clarify student queries regarding the questionnaire items.

Test administrators administered the student questionnaire successfully and delivered the field documents to the researcher. Student questionnaire data were entered into computers by the same data entry team that was involved in entering the Mathematics test. Similar to the procedures followed in entering the Mathematics test data, every morning data from the student questionnaire were entered into computers and every afternoon computerised data were screened for errors. The next section describes the various indices developed from the student questionnaire.

4.6.2.1.2 Simple Indices

Simple indices are constructed through arithmetical transformation or recoding of one or more items. Table 4.17 presents the list of simple indices constructed from the student questionnaire data (Full descriptions are available in Appendix 9)

Table 4. 17. Simple Indices Constructed from the Student Questionnaire

Simple Index	Description
Gender	Coded 1 for male and 2 for female
Age	Difference between the year of the survey and the year of students' birth
Parents' Occupational Status	Coded to ISCO (International Labour Office, 1990) and mapped to ISEI (Ganzeboom, Graaf, Treiman, & Leeuw, 1992; Ganzeboom & Treiman, 1996)

Parents' Educational Level	Coded to ISCED standards (OECD, 1999a; UNESCO, 1997)
Language Spoken at Home	Numbers were assigned to different languages or dialects reported by students
Expected Educational Level	Coded to ISCED standards (OECD, 1999a; UNESCO, 1997)
Grade Repetition	Numbers were assigned to responses
Late Arrival for School	Numbers were assigned to responses
Mathematics Homework	Numbers were assigned to responses
Time Spent on Mathematics Activities	Numbers were assigned to responses
Out-of-School-Time Lessons	Numbers were assigned to responses
Reasons for Choosing a School	Numbers were assigned to responses

4.6.2.1.3 Scale Indices Constructed from the Student Questionnaire

Nine multi-dimensional scale indices were constructed from the student questionnaire, and validated by performing confirmatory factor analyses (CFA). Table 4.18 presents the nine scale indices and a summary of the results of the CFA (Complete descriptions of all indices, including the factor loadings for individual measures, are available in Appendix 9).

Table 4. 18. Scale Indices with their Validity Indices as Constructed from CFA

Scale Index: ICT Confidence						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.08	0.96	0.96	0.05	ROUTCONF (a)	0.90	r_{ab}=0.75
				INTCONF (b)	0.90	r_{ac}=0.81
				HIGHCONF (c)	0.85	r_{bc}=0.85
Note 1: All validity criteria are met Note 2: Acronym: a=Confidence in routine ICT Tasks; b= Confidence in Internet Tasks; c=Confidence in High-Level ICT Tasks						

Scale Index: ICT Experience						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.05	0.90	0.89	0.08	INTUSE (a)	0.64	$r_{ab}=0.80$
				PRGUSE (b)	0.63	$r_{ac}=-0.15$
				ATTCOMP (c)	0.58	$r_{bc}=-0.50$
Note 1: a and b did not pass the discriminant validity criterion of validation. Note 2: a= Experience with Internet; b= Experience with programmes; c=Attitude towards ICT						
Scale Index: Learning Strategies						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.04	0.97	0.97	0.02	MEMOR (a)	0.55	$r_{ab}=0.97$
				ELAB (b)	0.66	$r_{ac}=0.96$
				CSTRAT (c)	0.64	$r_{bc}=0.92$
Note 1: a, b, and c did not pass the discriminant validity criterion of validation. Note 2: a= Memorization; b= Elaboration; c= Meta-cognition						
Scale Index: Motivations						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.06	0.97	0.96	0.02	INTMAT (a)	0.75	$r_{ab}=0.95$
				INSMOT (b)	0.73	
Note: a and b did not pass the discriminant validity criterion of validation. Note 2: a= Intrinsic motivation; b= Instrumental motivation						
Scale Index: Classroom Climate						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.05	0.93	0.91	0.04	TEACHSUP (a)	0.64	$r_{ab}=-0.27$

				DISCLIM (b)	0.64	
Note 1: All validity criteria are met Note 2: a= Teacher support; b= Disciplinary climate						
Scale Index: School Climate						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.05	0.93	0.91	0.03	ATSCHL (a)	0.73	r_{ab}=0.20
				STUREL (b)	0.62	r_{ac}=0.33
				BELONG (c)	0.56	r_{bc}=0.52
Note 1: b and c did not pass the discriminant validity criterion of validation. Note 2: a=Attitude towards school; b= Student and teacher relationship; c=Sense of belonging to school						
Scale Index: Preference for Learning Environment						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.09	0.92	0.89	0.03	COMPLRN (a)	0.72	r_{ab}=0.69
				COOPLRN (b)	0.68	
Note 1: a and b did not pass the discriminant validity criterion of validation. Note 2: a= Competitive learning environment; b= Cooperative learning environment						
Scale Index: Self-beliefs						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.42	0.96	0.96	0.04	MATHEFF (a)	0.75	r_{ab}=0.65
				SCMAT (b)	0.74	r_{ac}=-0.59
				ANXMAT (c)	0.73	r_{bc}=-0.94
Note 1: a and b did not pass the discriminant validity criterion of validation. Note 2: a=Self-efficacy in mathematics; b= Self-concept in mathematics; c= Anxiety in mathematics						

Scale Index: Household Possessions						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.46	0.94	0.95		COMPHOME (a)	0.87	r_{ab}=0.41
				CULTPOSS (b)	0.64	r_{ac}=0.63
				HERDES (c)	0.74	r_{bc}=0.85
Note 1: b and c did not pass the discriminant validity criterion of validation. Note 2: a=Computer facilities at home; b= Cultural possessions at home; c= Home educational resources						

4.6.2.1.4 Students' SES

An index of students' SES was derived as factor scores for the first principal component from the following three variables: (a) HISEI, (b) PARED, and (c) FWEALTH as explained in Appendix 9. Prior to factor scoring, the variables were standardized with a mean of zero and a standard deviation of one to assign a common metric. A principal components analysis was performed, by using SPSS version 19 (SPSS Inc, 2009), on the standardized variables to obtain factor scores for the first principal component. A similar method was used by the OECD (2005d, 2009b).

4.6.2.2 The Teacher Questionnaire

As mentioned in Chapter 3, a range of teacher-related educational effectiveness factors was identified at classroom level of the national educational assessment model. The teacher questionnaire was used to gather information about these factors.

Similar to the student questionnaire, various teacher questionnaires from the past studies that measured similar classroom-level factors were reviewed, with the view to adapting them for use in this study. Most of the items were adapted from NAEP and TIMSS. Some items were adapted from *Teaching and Learning International Survey* (TALIS). TALIS is a recent survey conducted by the OECD (2009a) in 24 countries to study school leadership roles and functions, teachers' professional development, appraisal and feedback, and beliefs and attitudes about teaching and their pedagogical practices. Table 4.19 shows the constructs used in the teacher questionnaire and their corresponding items, with the numbers in parentheses indicating the item numbers. The complete details of the teacher questionnaire items and their sources are provided in Appendix 10.

Table 4. 19. Constructs and Items used in the Teacher Questionnaire

Constructs	Questionnaire Items (Item Number)
Teacher Characteristics	Gender (1a-b); Age (2a-e); Academic degree (3a-d); Professional degree (4a-d); academic major (5a); Teaching experience (6a-e & 32a-e);
Instructional Activities	Use of time (7a-c); Teachers' professional collaboration (8a-d); Participation in PD (9a-s); PD needs (10a-i); Reasons for not participating in PD (11a-g); Activities in mathematics classroom (16a-z &17a-d)); Text book usage (15a-b); Teaching pedagogies (19a-e, 20a-e, & 21a-q); Readiness to teach (22a-g); Constraints on effective teaching (18a-n); Homework (23a-c,24a-e,25a-c, 26a-e); Tests (27a-e & 28a-c);
Appraisal and Feedback	Perceived priorities of appraisal and feedback (12a-p);
Beliefs and Attitudes	School safety (13a-d); School climate (14a-g); Self-efficacy (34a-i); Classroom climate (33a-e);
ICT	Calculators (29a-b & 30a-d); Computers (31a-d)

4.6.2.2.1 Field Administration

The 34-item teacher questionnaire was administered to the 60 teachers who were teaching mathematics to the Grade 10 Bhutanese students who participated in the Mathematics test. The teacher questionnaire was administered by the same test administrators who administered the Mathematics test. Like the student questionnaire, the teacher questionnaire was not pilot-tested with teachers.

Teachers' responses to the teacher questionnaire were coded and recorded. The same data entry team who entered student questionnaire data entered the teacher questionnaire responses into the Microsoft Excel spreadsheet. The data entry team entered data every morning, and every afternoon the team participated in screening these data for errors; with the errors corrected by referring to the relevant teacher questionnaire. The next section describes the types of indices developed from the teacher questionnaire.

4.6.2.2.2 Simple Indices

Simple indices were constructed from the teacher questionnaire by using the same procedure used in constructing simple indices from the student questionnaire. Table 4.20 shows the list of simple indices constructed from the teacher questionnaire together with brief procedural descriptions (Full descriptions are provided in Appendix 9).

Table 4. 20. Simple Indices Constructed from the Teacher Questionnaire

Simple Index	Description
Gender	Coded 1 for male and 2 for female
Age	Numbers were assigned to different age categories
Educational Attainment	Coded to ISCED (OECD, 1999a; UNESCO, 1997) standards
Academic Major	Responses were dichotomously coded
Teaching Experience	Numbers were assigned to different years of experience
Use of Time	Numbers were assigned to different amounts of time spent on various activities
Professional Development	Responses were dichotomously coded
Teacher Appraisal	Numbers were assigned to different levels of emphasis on different appraisal areas
Textbook Usage	Responses were dichotomously coded
Readiness to Teach	Numbers were assigned to different levels of readiness to teach different mathematics topics
Mathematics Homework	Numbers were assigned to the frequencies, duration, and types of homework
Assessment Practices	Numbers were assigned to the frequencies of the types of mathematics tests and questions used by teachers to assess students
Calculators and Computers	Numbers were assigned to teachers' usage of calculators and computers in mathematics class

4.6.2.2.3 Scale Indices Constructed from the Teacher Questionnaire

As was described in Section 4.6.2.1.3 on the student questionnaire, scale indices were derived from different items designed to measure a .latent construct of interest. The teacher questionnaire measured the following constructs related to teacher characteristics: teachers' professional collaboration; teachers' beliefs about teaching; classroom teaching practices (structured teaching, constructivist teaching, extended teaching, and use of reinforcements); teachers' self-efficacy; classroom climate; and school climate. Table 4.21 presents the scale indices and a summary of the results of the CFA (Complete descriptions, including the factor loadings for individual measures, are available in Appendix 9).

Table 4. 21. Scale Indices with their Validity Indices Constructed by Using CFA

Scale Index: Professional Collaboration (TPC)						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.00	1.00	1.05	0.70	TPC	0.74	
Note: All validity criteria are met						
Scale Index: Teachers' Beliefs about Teaching						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
*	*	*	*	TDTM (a)	0.58	
				TDCT (b)	0.36	
Note 1: * CFA could not be performed because of inadequate sample size as compared to the number of measures used in each dimension. Note 2: a=Index of beliefs about direct transmission teaching approach Note 3: b=Index of beliefs about constructivist teaching approach						
Scale Index: Structured Teaching (CTPST)						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.09	0.90	0.87	0.14	CTPST	0.75	
Note: Passed unidimensionality and convergent validity criteria of validation.						
Scale Index: Constructivist Teaching (CTPSO)						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.09	0.91	0.91	0.11	CTPSO	0.84	
Note: Passed unidimensionality and convergent validity criteria of validation.						
Scale Index: Extended Teaching (CTPEL)						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			

0.00				CTPEL	0.73	
Note: Passed unidimensionality and convergent validity criteria of validation.						
Scale Index: Use of Re-enforcements (TREM)						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.10	0.98	0.93	0.09	TREM	0.82	
Note: Passed unidimensionality and convergent validity criteria of validation.						
Scale Index: Factors Constraining Effective Teaching						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.21	0.68	0.62	0.21	TCONST (a)	0.75	r_{ab}=0.11
				TCONRT (b)	0.72	
Note 1: Passed all criteria of validity Note 2: TCONST=Index of student-related constraints Note 3: TCONRT=Index of resource-related constraints						
Scale Index: Teachers' Self-efficacy (TSE)						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.0	1.0	1.08	0.01	TSE	0.74	
Note: Passed unidimensionality and convergent validity criteria of validation.						
Scale Index: Classroom Climate (TCDC)						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.00	1.00	1.05	0.01	TCDC	0.69	
Note: Passed unidimensionality and convergent validity criteria of validation.						
Scale Index: School Climate (TSCHL)						
Goodness of Fit Indices				Dimension	Reliability	Correlation

RMSEA	CFI	NNFI	RMR			
0.08	0.96	0.98	0.01	TSCHL	0.84	
Note: Passed unidimensionality and convergent validity criteria of validation.						

4.6.2.3 The School Questionnaire

The school questionnaire sought to collect information about the various school effectiveness factors identified at the school level of the national educational assessment model proposed in Chapter 3. Similar to the student and the teacher questionnaires, different questionnaire items from the past surveys were reviewed, with the view to adapting them in the school questionnaire. Most of the items were adapted from PISA, TIMSS, and PIMRS. Table 4.22 shows the elements of these broad factors, with the numbers in the parentheses indicating the item numbers used in the school questionnaire. Complete information about the items and their sources are provided in Appendix 11.

Table 4. 22. Constructs and Items used in the School Questionnaire

Constructs	Questionnaire Items (Item Number)
School characteristics	School size (1a-b); Funding sources (2a-d); Assessment practices (7a-e and 8a-h); School admittance practices (6a-f); Monitoring strategies (12a-d); Enrichment activities (10a-e); Ability grouping (9a-d); School Autonomy (18a-f and 19a-f)
School resources	Quality of school resources (4h-i)); Quality of educational resources (4f-g, k-q); Teacher shortage (4a-c); Instructional time (3a-c); Quantity of teacher (11a-e); Availability of computers and the Internet (5a-f)
School climate	Factors affecting school climate (17a-l); Teacher consensus (13a-c,14a-c, and 15a-c); Teacher morale (16a-d)
Educational Leadership	Setting school goals (20a-e); Communicating school goals (21a-e); Supervision and evaluating instructions (22a-e); Coordinating school curriculum (23a-e); Monitoring student progress (24a-e); Protecting instructional time (25a-e); Maintaining high visibility (26a-e); Providing teacher incentives (27a-e); Providing professional development (28a-e); Providing incentives for learning (29a-e)
Parental Involvement	Parents' role (30a-e)

4.6.2.3.1 Field Administration

The school questionnaire was administered to the 60 school principals of the Grade 10 Bhutanese students who participated in the Mathematics test. The school questionnaire was administered by the same test administrators who administered the Mathematics test. Like the student and the teacher questionnaires, the school questionnaire was not pilot-tested with school principals.

Principals' responses to the school questionnaire were coded and recorded in line with the procedures followed for student and teacher questionnaires. The data entry team tabulated data in the morning and screened or errors in the afternoon. The following sections describe the types of indices developed from the school questionnaire.

4.6.2.3.2 Simple Indices

Similar to the process used in developing simple indices for student and teacher questionnaires; simple indices for the school questionnaire were developed by recoding or through arithmetic transformation of the variables. Table 4.23 shows the list of simple indices developed from the school questionnaire (Complete descriptions are available in Appendix 9).

Table 4. 23. Simple Indices Constructed from the School Questionnaire

Simple Index	Description
School Size	Total number of boys and girls
School Funding Sources	Numbers were assigned to different funding sources
School Admittance Policies	Numbers were assigned to different policy priorities followed by schools
Assessment Practices	Numbers were assigned to the frequencies and types of homework
Ability Grouping	Numbers were assigned to different types of ability groupings practised by schools
Mathematics Enrichment Activities	Responses were dichotomously coded for different enrichment activities
School Autonomy	Numbers were assigned to different types of school autonomy
Monitoring Strategies	Responses were dichotomously coded for different monitoring strategies

4.6.2.3.3 Scale Indices Constructed from the School Questionnaire

Scale indices were developed from the items that measured a construct of interest that had an underlying quantitative continuum. As was done with the student and the teacher questionnaires, items that measured the latent constructs of interest were validated by performing CFA. Table 4.24 presents the Scale indices and a summary of the results of the CFA (Details are available in Appendix 9).

Table 4. 24. Scale Indices with their Validity Indices Constructed by Using CFA

Scale Index: School Resources						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.00	1.00	1.00	0.03	Human Resources	0.65	
0.01	0.96	0.91	0.05	Material Resources	0.76	
Note: All validity criteria are met						
Scale Index: Teacher Morale						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.0				Teacher Morale	0.84	
Note: Passed unidimensionality and convergent validity criteria of validation.						
Scale Index: Teacher Consensus						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.0				Teacher Consensus	0.72	
Note: Passed unidimensionality and convergent validity criteria of validation.						
Scale Index: School Climate						
Goodness of Fit Indices				Dimension	Reliability	Correlation
RMSEA	CFI	NNFI	RMR			
0.16	0.88	0.80	0.04	Student-Related Factors	0.79	
0.00	1.00	1.02	0.03	Teacher-Related Factors	0.79	
Note: Passed unidimensionality and convergent validity criteria of validation.						
Scale Index: School Leadership						
Goodness of Fit Indices				Dimension	Reliability	Correlation

RMSEA	CFI	NNFI	RMR			
0.12	0.93	0.87	0.03	Goal Setting	0.74	
0.11	0.97	0.97	0.03	Communicating the Goal	0.81	
0.12	0.95	0.90	0.05	Supervising and Evaluating Instruction	0.79	
0.00	1.00	1.00	0.03	Coordinating School Curriculum	0.79	
0.11	0.96	0.93	0.04	Monitoring Student Progress	0.80	
0.04	0.91	0.82	0.05	Protecting Instructional Time	0.73	
0.16	0.91	0.82	0.05	Maintaining High Visibility	0.73	
0.02	1.00	0.99	0.03	Providing Teacher Incentives	0.82	
0.13	0.96	0.93	0.04	Providing Professional Development	0.84	
0.15	0.93	0.85	0.05	Providing Incentives for Learning	0.73	
Note: Passed unidimensionality and convergent validity criteria of validation.						

4.6.3 The Focus Group Interview Questionnaire

A two-hour focus group interview was conducted with a three-member focus group. The group was introduced to the focus group research agenda by presenting a short abstract, which was then followed by group discussion. The discussion was guided by the 16 questions. The complete focus group interview guide is available in Appendix 12. The researcher performed the role of the group moderator. The discussions were audio-recorded on a digital voice recorder. The audio record was transcribed by the researcher on the same day as the focus group interview, and was made ready for further analyses.

4.7 Chapter Summary

The chapter began by briefly revisiting the research aim and outcomes as a prelude to selecting appropriate research paradigms for the thesis. Pragmatism was chosen as the appropriate paradigm for this study. This paradigm naturally qualified mixed methods as the research method for this study. A cross-sectional survey was used as part of the quantitative method and a focus group interview was used as part of the qualitative method. The use of both quantitative and qualitative methods qualified the overall research method as the mixed methods research within the pragmatist paradigm, with the quantitative method as the dominant one.

The sampling design for the study was described. Drawing on the pros and cons of the various factors, which influence sampling, a two-stage-stratified cluster sampling design was used for sampling participants for the cross-sectional survey. A purposive sampling design was used for sampling participants for the focus group interview.

The items used in the research instruments were adapted from the past studies such as PISA, TIMSS, NAEP, TALIS, and PIMRS. The instruments, namely, the Mathematics test, the student questionnaire, the teacher questionnaire, and the school questionnaire were administered in schools by trained test administrators. Data collected from these instruments were screened for errors using Microsoft Excel spreadsheet and SPSS. Psychometric properties like reliability and other validity statistics of the Mathematics test and the scale indices were reported.

4.8 Chapter Conclusion

This chapter presented the research designs, the sample designs, and the research instruments of the study. Pragmatism, mixed-methods, cross-sectional survey, and focus group interviews are discussed as parameters of the research design of the study. A two-staged-stratified cluster sampling method and a purposive sampling method are discussed as the sampling designs of the study, with the former for sampling the participants for the cross-sectional survey and the latter for sampling the participants for the focus group interviews. A Mathematics test, a student questionnaire, a teacher questionnaire, and a school questionnaire are selected as the instruments for collecting data from the participants in the cross-sectional survey, and a focus group interview is mounted as the instrument for collecting data from its Ministry participants. Finally, the chapter presented several indices and their validity statistics, including the psychometric properties of the Mathematics test.

Chapter 5

ANALYSES OF THE MATHEMATICS TEST

5.1 Introduction

The purpose of this chapter is to report the analyses of data from the Mathematics test. The analyses were carried out in line with the research outcomes 1, 2, 3, 5, 8, and 9 of generating knowledge about: the profiles of students about their comprehension of school curricular content; student outcomes at school, district, regional, and national levels; about the skills that Bhutanese students need to adapt to rapid societal and technological change; the equity in, and accessibility to, educational resources and provision; and the standard of the Bhutanese school education system as compared to the standard of other countries. Specifically, the results of the analyses in this chapter provide responses to the following research questions from Section 4.3.2.1:

OuL 1 What is the profile of Bhutanese Grade 10 students' mathematical knowledge and skills at the end of their basic education?(Sections 5.2 and 5.3)

StL 9 How well do Bhutanese Grade 10 students rate the higher-order thinking skills scale? (Section 5.4)

StL 12 Do Bhutanese Grade 10 students have mathematical knowledge and skills to adapt to rapid societal and technological change? (Section 5.3 and 5.4)

StL 13 How does Bhutanese student achievement vary across school, district, regional, national levels? (Sections 5.5, 5.6, and 5.7)

StL 14 How does Bhutanese Grade 10 student achievement in Mathematics compare with international students? (Section 5.8)

Data from the test were analysed by observing the following analytical framework: (a) describing data analysis procedures, (b) presenting statistical indices, (c) relating research outcomes to statistical indices, and (d) informing policy perspectives. Guided by this analytical framework, the purposes of this chapter are to: profile students' mathematical knowledge and skills; profile students' thinking skills; compare schools, school districts, and locale; set international benchmarks; and discuss policy implications.

5.2 A Profile of Students' Mathematical Knowledge and Skills

Profiling students' mathematical knowledge and skills in terms of individual mathematics topics or mathematics curriculum strands has the advantage of providing diagnostic information about students' understanding of the individual mathematics topics or curriculum strands. This section answers the research question, "What is the profile of Bhutanese Grade 10 students' mathematical knowledge and skills at the end of their basic education?

It is possible to profile students' mathematical knowledge and skills by using a test specification. As mentioned in Chapter 4, two test specifications were developed for the Mathematics test—one in terms of the PISA Mathematical Literacy Framework as shown in Table 4.3 and another one in terms of the Bhutanese Grade 10 mathematics curriculum as shown in Table 4.5. Further, the profiles can be interpreted with the PISA Mathematical Proficiency Scale and Levels presented in Table 4.15.

First, Table 5.1 shows the profile of Grade 10 Bhutanese students' mathematical knowledge and skills with reference to the PISA Mathematical Literacy Framework.

Table 5. 1. Grade 10 Bhutanese Students' Mean Mathematics Test Scores on the Four Domains of the PISA Mathematical Literacy Framework

Domain	Mean	S.E	Location on the PISA Mathematics Proficiency Scale
Shape & Space	355	3.04	Below Level 1
Quantity	450	2.53	Level 2
Change & Relationship	414	2.21	Level 1
Uncertainty	385	2.47	Level 1

Table 5.1 indicates that, on average, Grade 10 Bhutanese students' scored below, or at, Level 1 of the PISA Mathematics Proficiency Scale in the domains of Space and Shape, Change and relationship, and Uncertainty of the PISA Mathematical Literacy Framework, and at Level 2 of the PISA Mathematics Proficiency Scale in the domain of Quantity of the PISA Mathematical Literacy Framework.

Second, the mean performance scores of the students in the individual strands of the Bhutanese Grade 10 mathematics curriculum are computed. Table 5.2 shows the result of the computation.

Table 5. 2. Bhutanese Grade 10 Students' Mean Mathematics Test Scores on the Six Strands of the Bhutanese Grade 10 Mathematics Curriculum

Strands	Mean	S.E	Location on the PISA Mathematics Proficiency Scale
Numbers	465	4.73	Level 2

Operations	385	7.15	Level 1
Pattern	409	4.80	Level 1
Measurement	408	3.74	Level 1
Geometry	465	4.41	Level 2
Data Management & Probability	383	2.81	Level 1

Table 5.2, indicates that Grade 10 Bhutanese students are able to demonstrate the mathematical knowledge and skills described in Level 2 of the PISA Mathematics Proficiency Scale in the strands of Numbers and Geometry of the Grade 10 mathematics curriculum. Table 5.2 also shows that Grade 10 Bhutanese students are able to demonstrate mathematical knowledge and skills described in Level 1 of the PISA Mathematics Proficiency Scale in the strands of Operations, Pattern, Measurement, and Data Management and Probability of the Grade 10 mathematics curriculum.

Discussion

By profiling students' knowledge and skills by subject content or standards, it is possible for the parents, the teachers, and the schools to understand students' strengths and weaknesses in terms of mathematics curriculum content or standards. Such understanding provides the parents, the teachers, and the schools with critical insights into students' subject knowledge and skills with which they can offer strategic remedial lessons or other personalised support and guidance to the students (OECD, 2004a, 2005a, 2007a).

The profiles of students' mathematical knowledge and skills in terms of the four domains of the PISA 2003 Mathematical Literacy Framework and the Bhutanese Grade 10 Mathematics Curriculum revealed that on average Grade 10 student performance corresponded to a maximum of Level 2 of the PISA Mathematics Proficiency Scale and at a minimum of below Level 1 of the PISA Mathematics Proficiency Scale. This indicates great scope for improvement in teaching pedagogies and development of educational interventions. This is all the more plausible because of a widely held view that teachers' subject content and pedagogical knowledge and skills, and students' characteristics, and student achievement are closely related (NCTM, 2000).

5.3 Percentages of Students at each Level of the PISA Mathematics Proficiency Scale

To provide further insight into the profiles of students' mathematical knowledge and skills, the percentage of students at each level of the PISA Mathematics Proficiency Scale was calculated. Table 5.3 presents the result of the computation. Table 5.3 also suggests an answer to the research question, “Do Bhutanese students Grade 10 students have mathematical knowledge and skills to adapt to rapid societal and technological change?”

Table 5. 3. Percentages of Bhutanese Students at Individual PISA Mathematics Proficiency Levels

Proficiency Level	Percentage	SE
Below 1	27.03	2.03
1	35.16	1.67
2	26.62	1.44
3	8.092	1.26
4	2.7	0.65
5	0.30	0.14
6	0.10	0.08

Table 5.3 indicates that approximately a quarter of Grade 10 Bhutanese students scored below Level 1 of the PISA Mathematics Proficiency Scale. Table 5.3 also shows that the majority of students performed at Level 1 and Level 2 of the PISA Mathematics Proficiency Scale. Approximately one-tenth of the students performed at Level 3 or higher on the PISA Mathematics Proficiency Scale.

Discussion

The OECD (2005a) noted that one school year equals an average of 41 score points on the PISA Mathematics Proficiency Scale. This indicates that the students who scored below Level 1 of the PISA Mathematics Proficiency Scale had very limited mathematical knowledge and skills expected of a Grade 10 Bhutanese student. In other words, the students who failed to attain Level 1 of the PISA Mathematics Proficiency Scale had difficulty in solving mathematical tasks that required them to use factual knowledge and routine procedures. The OECD considers the students scoring below Level 1 of the PISA Mathematics Proficiency Scale as being at risk of not achieving the mathematical knowledge and skills required for them to participate fully in society beyond school (Thomson & Bortoli, 2008). Therefore, Grade 10 students who scored below Level 1 of the PISA Mathematics Proficiency Scale do not have the mathematical knowledge and skills necessary for successful adaptation to rapid societal and technological change.

5.4 A Profile of Students' Thinking Skills

As described in Chapters 2 and 4, students' thinking skills are grouped into three broad competency clusters, namely, reproduction, connections, and reflection. Table 5.4 presents the students' mean performance scores on these three broad competency clusters.

Table 5. 4. Mean Scores on the Three Competency Clusters of the Mathematics Test

Competency Clusters			Reproduction	Connections	Reflection
	Mean		463	365	353
		SE	4.72	3.22	3.73
Reproduction	463	4.72	0	^	^
Connections	365	3.22	v	0	v
Reflection	353	3.73	v	^	0

Note: 1. The table is read across the row for a cluster to compare with the clusters listed along the top of the table.
2. v denotes less than and ^ denotes greater than.
3. The standard errors include linking error as well.

Multiple comparisons of the students' mean scores on the competency clusters, using the Bonferroni correction, showed that the students' mean performance scores on the competency clusters were statistically significantly different from each other, $p<0.05$. This indicates that students differ in their range of thinking skills, and that some of them are capable of using higher-order thinking skills to solve complex mathematical problems.

Discussion

Higher-order thinking skills are widely accepted as necessary skills to pursue life-long learning, which in itself is considered as a means to cope with rapid societal change (Mullis, et al., 2005; OECD, 2007a). As shown in Table 5.4, this study grouped students' thinking skills into three competency clusters: reproduction, connections, and reflection. These clusters were described in detail in Chapter 2. The pattern in the mean scores across the competency clusters shows that the students are not as good at applying higher-order thinking skills (connections and reflection clusters) as they are at applying lower-order thinking skills (reproduction cluster). The mean score on the test items from the reproduction cluster is higher than the mean score on the test items from the connections cluster and the mean score on the items from the connections cluster is higher than the mean score of the items from the reflection cluster. Because the reproduction cluster has all the aspects of rote learning, this finding suggests that the students use rote learning as the principal strategy for learning mathematics. This finding also corroborates the findings of a couple of studies commissioned by the Royal Education Council of Bhutan (Educational Initiatives, 2009; iDiscoveri Education & Royal Education Council, 2009).

5.5 Mean Mathematics Test Scores of Students at National, District, and School Levels

Student achievement when disaggregated at district and school levels yields information that is more relevant to stakeholders than the information conveyed by student achievement that is aggregated only at the national level. Therefore, this section presents the mean performance scores of students at national, district, and school levels.

First, the national mean Mathematics test score of Bhutanese Grade 10 students was 361 (S.E.= 4.1), resulting in the 95% confidence limits for the population mean being in the range of ±8.04 test score points. This value was just within the prior specification of the 95% confidence limits of ±10 test score points as mentioned in Section 4.4.6 of Chapter 4, indicating that the sample size of 1,500 students had satisfied the prior sampling error specification.

Second, the mean Mathematics test scores of the 19 districts of Bhutan are computed as shown in Table 5.5. The one-way between-group analysis of variance showed a statistically significant difference in the mean Mathematics test scores of the 19 districts, $p<0.05, F(18,1495) = 3.18, p = 0.00$. Post-hoc comparisons using the Bonferroni correction showed that the mean performance scores of some districts were statistically significantly different from other districts, $p<0.05$. However, most of the districts did not have statistically significant differences in their mean performance scores with other districts.

Table 5. 5. Comparison of the Mean Mathematics Test Scores of 19 Districts

			A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**
Mean			*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
		SE	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
A	**389**	8.0	0	•	v	•	•	•	•	•	•	•	•	•	•	•	•	•	v	•	•
B	**412**	13.2	•	0	v	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
C	**468**	5.8	^	^	0	^	^	^	^	^	^	^	^	^	^	^	^	^	^	^	^
D	**415**	5.8	•	•	v	0	•	•	•	•	•	•	•	•	•	•	•	^	•	•	•
E	383	18.7	•	•	v	•	0	•	•	•	•	•	•	•	•	•	•	•	v	•	v
F	396	38.0	•	•	v	•	•	0	•	•	•	•	•	•	•	•	•	•	v	•	•
G	396	19.8	•	•	v	•	•	•	0	•	•	•	•	•	•	•	•	•	v	•	•

H	**413**	13.5	•	•	v	•	•	•	•	0	•	•	•	•	•	•	•	•	v	•	•
I	**396**	9.2	•	•	v	•	•	•	•	•	0	•	•	•	•	•	•	•	v	•	•
J	398	26.8	•	•	v	•	•	•	•	•	•	0	•	•	•	•	•	•	v	•	•
K	**410**	21.4	•	•	v	•	•	•	•	•	•	•	0	•	•	•	•	•	v	•	•
L	**395**	14.9	•	•	v	•	•	•	•	•	•	•	•	0	•	•	•	•	v	•	•
M	**400**	6.7	•	•	v	•	•	•	•	•	•	•	•	•	0	•	•	•	v	•	•
N	395	17.4	•	•	v	•	•	•	•	•	•	•	•	•	•	0	•	•	v	•	•
O	**407**	19.6	•	•	v	•	•	•	•	•	•	•	•	•	•	•	0	•	v	•	•
P	**377**	5.8	•	•	v	v	•	•	•	•	•	•	•	•	•	•	•	0	v	•	v
Q	**433**	5.8	^	•	•	•	^	^	^	•	^		•	^	•	^	•	^	0	•	•
R	**406**	10.7	•	•	v	•	•	•	•	•	•	•	•	•	•	•	•	•	•	0	•
S	**420**	15.6	•	•	v	•	^	•	•	•	•	•	•	•	•	•	•	^	•	•	0

Note:1. The table is read across the row for a district to compare its performance with the districts identified along the top of the table.
2. v denotes less than, ^ denotes greater than, * refers to the means and standard errors shown in the columns under 'Mean' and 'SE' for the corresponding districts, and • denotes non-statistically significant difference.
3. The standard errors include the link errors as well.
4. Bold numbers indicate a statistically significant difference from the national mean performance scores at p<0.05.

Third, the mean performance scores of the 60 schools were calculated. Table 5.6 presents a portion of the result of the post-hoc test, for 12 schools. The complete result is available in Appendix 13. The one-way between-group analysis of variance showed a statistically significant difference at the p<0.05 level in the mean Mathematics test scores of the 60 schools; $F(59{,}1495) = 3.8$, $p = 0.00$. Post-hoc comparisons of schools using the Bonferroni correction showed that the mean scores of some schools were statistically significantly different from that of other schools, p<0.05. However, most of the schools did not have statistically significant differences in their mean performance scores with other schools.

Table 5. 6. Comparison of the Mean Mathematics Test Scores of 60 Schools

			1	2	3	4	5	6	7	8	9	12	...
Mean			404	395	387	433	439	388	470	381	365	391	
		SE	14.6	15.1	11.6	13.6	13.9	11.6	12.1	16.2	15.1	17.4	
1	**404**	14.6	0	•	•	•	•	•	•	•	•	•	
2	**395**	15.1	•	0	•	•	•	•	•	•	•	•	
3	**387**	11.6	•	•	0	•	•	•	**v**	•	•	•	
4	**433**	13.6	•	•	•	0	•	•	•	•	•	•	
5	**439**	13.9	•	•	•	•	0	•	•	•	•	•	
6	**388**	11.6	•	•	•	•	•	0	**v**	•	•	•	
7	**470**	12.1	•	•	^	•	•	^	0	^	^	•	
8	381	16.2	•	•	•	•	•	•	**v**	0	•	•	
9	365	15.1	•	•	•	•	•	•	**v**	•	0	•	
12	391	17.4	•	•	•	•	•	•	•	•	•	0	
. . .													

Note: 1. The table is read across the row for a school to compare its performance with the schools identified along the top of the table.

2. v denotes less than, ^ denotes greater than, and • denotes non-statistically significant difference.
3. The standard errors include the link errors as well.
4. Bold numbers indicate a statistically significant difference from the national mean performance scores at $p<0.05$.

Discussion

The mean Mathematics test scores of students at national, district, and school levels have the potential to encourage schools and districts to exchange their expertise and learn from one another's strengths and weaknesses. Such potential is even greater when the difference in mean performance scores of schools is statistically significant. The results of the analyses showed that schools and districts differed in the mean Mathematics test scores of their students. This observation also supports Barton's (2002) statement that student achievement varies across population sub-groups, schools, districts, and regions.

The mean performance scores of some schools differed significantly from those of other schools. This shows that there is an opportunity for the schools to learn from their comparative strengths and weaknesses. As presented by Barber and Mourshed (2007), schools can develop programs to visit or interact with both high-performing and low-performing schools to study some of their characteristics and the tools they use to improve student outcomes. Similarly, student performance differed across districts, making it possible for District Education Officers to learn from one another. The exchange of ideas between schools, between districts, or between schools and districts, can drive profitable educational interventions (Mullis, et al., 2004; OECD, 2007a; Weller, 1996). In addition, differences in the performance of schools has the potential to facilitate school choice by parents which can increase parental participation in school activities and influence schools' decision-making processes, resulting in improved student achievement (Mullis, et al., 2004; OECD, 2007a).

5.6 Performance of Students by their Schools' Locale

The policy of the Ministry of Education clearly identifies schools into four categories by using schools' locale in terms of setting. The four categories, as discussed in Chapter 4, are (a) urban, (b) semi-urban, (c) semi-remote, and (d) remote. The mean performance scores of students by their schools' locale in relation to setting were computed. Table 5.7 shows the results of the calculation.

Table 5. 7. Comparison of Students Mean Mathematics Test Scores by their Schools' Locale (Category)

Locale (Category)			Urban	Semi-Urban	Semi-Remote	Remote
	Mean		365	353	318	337
		SE	4.39	9.25	13.58	6.07
Urban	365	4.39	0	12	**47**	**28**
Semi-Urban	353	9.25	-12	0	**35**	16
Semi-Remote	318	13.58	**-47**	**-35**	0	-19
Rural Remote	337	6.07	**-28**	-16	19	0

Note: 1.The standard errors include the link errors as well.
2. The significant differences are shown in bold, $p<0.05$.

Table 5.7 shows that the mean Mathematics test score of the students in urban schools is higher than those of the students in semi-remote and remote schools. Likewise, the mean score of the students in semi-urban schools is statistically significantly higher than the mean score of the students in semi-remote schools.

Discussion

The performances of Grade 10 Bhutanese students differed significantly in terms of their schools' locale. As expected, students in urban schools outperformed the students in semi-urban schools (although not statistically significant), and the students in semi-urban schools outperformed the students in semi-remote and remote schools. This pattern is consistent with the findings of similar studies commissioned by the Royal Education Council of Bhutan (Educational Initiatives, 2009; iDiscoveri Education & Royal Education Council, 2009). The findings signal the need for relevant educational interventions aimed at improving student outcomes in semi-rural and rural schools as compared to urban and semi-urban schools.

5.7 The Between- and Within-School Variances

The between- and the within-school variances in performance scores of Grade 10 Bhutanese students were computed by using the VARCOMP procedure available from SPSS version 19 (SPSS Inc, 2009), The within-school variance (37198.20) is comparatively larger than the between-school variance (4597.53), with 11% of the variance explained by the between-school variance.

Discussion

Students' test scores are mostly used as measures of the effectiveness of an education system. However, the school mean performance scores alone do not present a complete picture of the effectiveness of an education system. As described in Chapter 3, an effective education system should have a small difference in performance between schools.

Eleven percent of the variance in students' performance is explained by the difference between schools, and the rest is explained by the variance within schools. This finding is similar to the findings in other countries where a comprehensive school system, as in Bhutan, is practised (OECD, 2004a). This between-school variance, which is about one-third of the OECD's average (OECD, 2004a, p. 162), is attributable to the lack of streaming or tracking practices, the absence of parental choice of schools, and the use of common school curriculum in the Bhutanese education system. In addition, school leadership practices, teacher characteristics, and SES are other possible factors that might have contributed to the 11% between school variance. The remaining within school variance, that is, 89% may be attributable to teacher and student characteristics and classroom climate (Hattie, 2009). The implication from the finding is that Bhutan is well on track to achieving equity in educational opportunities as far as the school education system is concerned.

5.8 International Benchmarks for the Bhutanese Education System

The national mean of the Mathematics test score of Grade 10 Bhutanese students can be compared with the national mean scores of the countries that participated in PISA 2003, because the

Mathematics test and the PISA 2003 Mathematical Literacy Test are linked. The procedures applied in linking the two tests are described in Chapter 4 in Section 4.6.1.7. The linkage between the two tests makes it possible to set international benchmarks for the Bhutanese education system. Using the national mean scores of the countries that participated in PISA 2003 and the national mean score of Grade 10 Bhutanese students on the Mathematics test, a league table is thus developed as shown in Table 5.8.

Table 5. 8. Bhutan Benchmarked with the Countries that Participated in PISA 2003

Country	Mean	SE
Hong Kong-China	550	4.5
Finland	544	1.9
Korea	542	3.2
Netherlands	538	3.1
Liechtenstein	536	4.1
Japan	534	4
Canada	532	1.8
Belgium	529	2.3
Switzerland	527	3.4
Macao-China	527	2.9
Australia	524	2.1
New Zealand	523	2.3
Czech Republic	516	3.5
Iceland	515	1.4
Denmark	514	2.7
France	511	2.5
Sweden	509	2.6
Austria	506	3.3
Germany	503	3.3
Ireland	503	2.4

OECD average	***500***	***0.6***
Slovak Republic	498	3.3
Norway	495	2.4
Luxembourg	493	1
Hungary	490	2.8
Poland	490	2.5
Spain	485	2.4
United States	483	2.9
Latvia	483	3.7
Russian Federation	468	4.2
Italy	466	3.1
Portugal	466	3.4
Greece	445	3.9
Serbia	437	3.8
Turkey	423	6.7
Uruguay	422	3.3
Thailand	417	3.0
Mexico	385	3.6
BHUTAN	361	4.1
Indonesia	360	3.9
Tunisia	359	
Brazil	356	4.8

Note: 1. The national mean score data of the countries that participated in PISA 2003 are adapted from the OECD (2004a, p. 358).

Table 5.8 shows that Grade 10 Bhutanese students' mean score is greater than the mean scores of Indonesia (360), Tunisia (359), and Brazil (356).

Percentages at different levels of the PISA Mathematics Proficiency Scale also provide international benchmarks. Figure 5.1 shows the percentages of students at each level of the PISA Mathematics

Proficiency Scale of Grade 10 Bhutanese students and students in the countries that participated in PISA 2003. Figure 5.1 is constructed with data from PISA 2003 (OECD, 2004a, p. 354) and the Mathematics test. This figure displays a more detailed but similar benchmark for Bhutan.

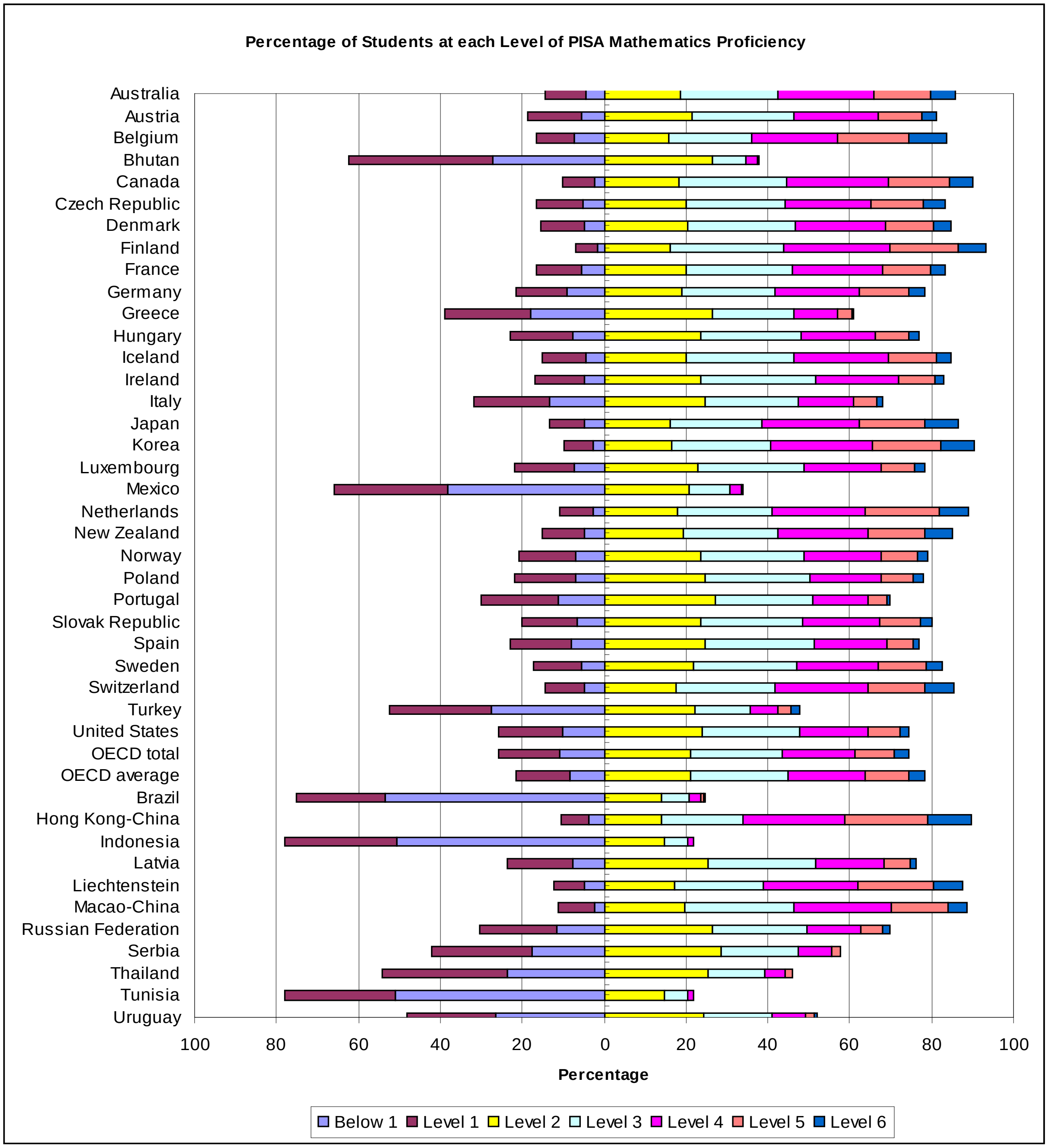


Figure 5. 1. Percentages of Students at each Level of the PISA Mathematics Proficiency Scale with Bhutan Included

On average, only about a third of students across the OECD countries attained level five and six of the PISA Mathematics Proficiency Scale (OECD, 2005a), indicating the challenge of achieving high national scores in this subject as is the case with Grade 10 Bhutanese students.

Discussion

Published perspectives and opinions of education systems are developed on the basis of student performance as implied by popular international assessment programs, such as PISA, TIMSS, and NAEP; although the validity of such an approach to judging an education system is subject to on-going debate. This study used the mean scores and the performance levels of Bhutanese students on the Mathematics test to locate the position of the Bhutanese education system among the education systems of the countries that participated in PISA 2003 (OECD, 2004a). First, the mean score of Bhutanese students on the Mathematics test corresponded to the 39th position on the list of 42 countries that participated in the PISA 2003—preceding Indonesia, Tunisia, and Brazil. Second, the distribution of the scores of students from Bhutan on the PISA Mathematical Proficiency Scale is similar to the students from Mexico, Indonesia, Tunisia, and Brazil. The majority of the students in these countries performed below Level 1 or at Level 1 of the PISA Mathematical Proficiency Scale. These findings indicate that Bhutan has a large range of opportunities to learn from the education systems of the countries that participated in PISA (OECD, 2004a, 2007a, 2010), both from high- and low-performing countries.

However, a word of caution has to be emphasized while comparing the scores from two different tests. As emphasized by Johnson and Owen (1998), comparisons of scores from two linked tests should be viewed as estimates, and not as substituting one test for the other. Notwithstanding the cautionary note, such estimates provide guidelines for between-country fertilization of innovative approaches in school education. For example, countries with similar characteristics can learn from their weaknesses and strengths through exchange of knowledge and skills related to school effectiveness.

5.9 Summary and Implications for Policy

The results from the analyses have a range of implications for policy. The majority of students performed at or below Level 1 of the PISA Mathematics Proficiency Scale, indicating that they were more proficient in recalling facts and using procedural knowledge than in applying higher-level knowledge and skills. Therefore, the Bhutanese education system needs policy interventions to develop higher-level mathematical literacy in its students. Further, percentages of students at different levels of the PISA Mathematics Proficiency Scale showed that the students differed widely in their performance, indicating the need for a range of teaching pedagogies and school curricula that are capable of engaging students of different abilities. A lack of educational opportunities capable of engaging students of different abilities may promote a pyramidal pattern in enrolments from kindergarten to university, with many low-ability students dropping out of the system before completing their university-level education. Some of the implications from the pyramidal pattern are that many students leave school with inadequate knowledge and skills necessary for gainful participation in the nation's economy or for pursuing life-long learning. It was also shown that 11% of the variance in student performance was accounted for by the between-school differences, suggesting that schools in Bhutan had similar characteristics. Such a scenario could have been an indication of the Bhutanese education system having achieved equity and accessibility in educational opportunities, had its overall student achievement level been at the higher end of the PISA Mathematics Proficiency Scale. Therefore, it is desirable for policy makers to focus on educational interventions that will raise the overall student performance level. This may be achieved by designing interventions at systemic level (e.g., teacher education). Finally, students in urban schools outperformed students in other locales on the Mathematics test, suggesting the contribution of schools' locale to the between-school variance. This indicates the need for the Ministry of Education to develop educational interventions aimed at providing equitable access to educational opportunities for all students and schools.

As with the implications for policies from the analyses of the Mathematics test data, the methodological aspects of the analyses also have potential to guide the current assessment practices in the Bhutanese education system. The method of linking different tests has the potential to help the Bhutan Board of Examinations conduct trend studies in student achievement by using its annual examination data. The method of using a proficiency scale has the potential to help the Ministry of Education of the Royal Government of Bhutan relate its curriculum standards to students' test scores. Such a relation would provide systemic feedback to stakeholders (e.g., policy-makers, schools, teachers, parents, students). Likewise, the method of profiling students' knowledge and skills has the potential to help the Bhutan Board of Examinations profile students' knowledge and skills in different school subjects by using its national examination data. The method followed in the use of proficiency levels has the potential to assist the Ministry of Education of the Royal Government of Bhutan in setting benchmarks, organising inter-school exchange of educational expertise, and developing professional development programmes. Further, the method of profiling students' thinking skills has the prospect of helping the Ministry of Education of the Royal Government of Bhutan in devising teaching and learning pedagogies and school curriculum that provides students with opportunities for developing higher-order thinking skills. The method of disaggregating students' test scores at school, district, and national levels has the potential to help the Ministry of Education of the Royal Government of Bhutan benchmark its schools with one another, which would enable the schools to learn from their strengths and weaknesses through various educational exchange-programmes. The method of comparing schools in terms of their locale has the potential to help the Bhutan Board of Examinations generate similar information, which would help the Ministry of Education of the Royal Government of Bhutan allocate educational resources where they are most needed. Finally, the method of processing within- and between-school variances can help the Bhutan Board of Examinations in generating similar information from its examination data, which would offer policy-makers a broad overview of the systemic potential and challenges.

5.10 Chapter Conclusions

This chapter provided diverse insights into the Bhutanese education system, and revealed a range of implications for educational policies. Grade 10 Bhutanese students' mathematical knowledge and skills ranged from below Level 1 to Level 6 of the PISA Mathematics Proficiency Scale, and the percentages of the students at each level of the PISA Mathematics Proficiency Scale ranged from 27% below Level 1 to 0.1% at Level 6. These indicate great scope for improvement in teaching pedagogies and development in educational interventions. Furthermore, Grade 10 Bhutanese students performed better on mathematical problems that required lower-order thinking skills than the problems that required higher-order thinking skills. Student achievement also differed between districts, between schools, and between schools' locale, indicating the need for educational interventions capable of promoting equitable access to educational opportunities for schools. This finding also shows that school districts, and schools individually, have much to learn from their differences through programmes such as exchange visits. Last but not least, the international benchmarks lined up against the Bhutanese education system showed that Bhutan had much to learn from the countries that participated in PISA. Bhutan could look for innovative educational models in the top-performing countries like Hong Kong-China, Finland, Korea, and the Netherlands. Bhutan could also look at alternative experiences in some low-performing countries like Brazil, Tunisia, and Indonesia.

Chapter 6

ANALYSES OF THE STUDENT QUESTIONNAIRE

6.1 Introduction

The purpose of this chapter is to report the analyses of data from the student questionnaire. The purposes of the analyses are to address the research outcomes 2, 5, and 8 of generating knowledge about: the factors related to effective schooling and their effects on student outcomes; the preparedness Bhutanese students to meet the challenges of the future; and the equity in, and accessibility to, educational resources and provision. Specifically, this chapter seeks to answer the following research questions from Section 4.3.2.1:

StL 1 How well do gender, age, and SES relate to student achievement? (Sections 6.2, 6.3 and 6.4)

StL 2 How well does motivation relate to student achievement? (Section 6.5)

StL 3 How well do self-beliefs relate to student achievement? (Section 6.6)

StL 4 How well do meta-cognitive skills relate to student achievement? (Section 6.7)

StL 5 What are students' learning preferences? (Section 6.8)

StL 6 How well does ICT relate to student achievement? (Section 6.9)

StL 7 What do students think of their classroom and school climates? Sections 6.10 and 6.11)

StL 8 How well does homework relate to student achievement? (Section 6.12)

StL 10 Are Bhutanese students engaged in their schools? (Sections 6.6, 6.10, and 6.11)

StL 11 Are Bhutanese students prepared for the challenges of the future? (Section 6.7)

OuL 2 What are the affective responses of Bhutanese students to their basic education? (Sections 6.6 and 6.11)

The analyses were done in line with the analytical framework followed in Chapter 5, and the educational effectiveness factors were analysed in the order in which they had been discussed in Chapter 3.

6.2 Students' Gender and Mathematics Achievement

The difference in the average Mathematics test scores of the female (mean=349.86, SE=5.21) and the male (mean=370.57, SE=4.31) students was statistically significant, with the male students, on average, scoring higher than the female students; *diff*=20.71, 95% CI [10.92, 30.50].

Discussion

This finding corroborates the findings reported in similar studies conducted in Bhutan (Bhutan Board of Examinations, 2008; Educational Initiatives, 2009). While gender-differentiated performance in mathematics in Bhutan has been reported, educational interventions to bridge the gap have hardly been implemented. The causes for the gender-differentiated performance are attributable to peer relationships (Crosnoe, et al., 2008), hetero-femininities (Archer, Halsall, & Hollingworth, 2007), comprehensive or co-education system (Malacova, 2007; Younger & Warrington, 2006, 2007), gendered academic cultures (Houtte, 2004a, 2004b), and gendered achievement goals (Graham, et al., 2008). Given that gender has been reported as a factor in student learning in a number of studies, an independent study on gender and student learning is desirable for Bhutan.

6.3 Students' Age and Mathematics Achievement

Age as a factor of student performance on the Mathematics test was analysed using regression and correlational analyses.

The Pearson correlation coefficient between students' age and their performance on the Mathematics test was significantly different from zero, r=-0.27, 95% CI [-0.33, -0.21], showing that age was inversely related to performance on the Mathematics test.

A linear regression analysis revealed that the age explained 7.3% of variance in the Mathematics test score, $R^2 = 0.073$, 95% CI [0.04, 0.11]. A unit increase in age corresponded to a statistically significant decrease of 11.66 on the Mathematics test scores, B=-11.66, 95% CI [-14.63, -8.70].

The age at which the performance on the Mathematics test was optimal ranged from 15 to 17 years. Given that the official school enrolment age of Bhutanese children was six years, a majority of the Grade 10 students should be in the optimal age range. However, the average age of the students in Grade 10 was 18.55 years, 95% CI [18.40, 18.70]; indicating grade retention or late enrolment. The students were asked whether they repeated one or more grades. Thirty percent of students reported repeating a grade at primary school, 95% CI [26.60, 32.54]. Another three percent reported repeating two or more grades at primary school, 95% CI [2.03, 3.93].

Discussion

Younger students outperformed older students. Given that the official enrolment age was six years in Bhutan, the expected average age of Grade 10 students should be in the range of 15 to 17 years. However, the actual average age was 18.55 years, indicating grade retention or late enrolment. Analyses of data on grade retention showed that over thirty percent of the students repeated at least one grade before reaching Grade 10. Grade retention and academic redshirting do not improve student outcomes (Braymen & Piersel, 1987; Cameron & Wilson, 1990; Jimerson, 2001; Jimerson, et al., 2002; Jimerson, et al., 2006). On the contrary, grade retention is likely to put pressure on educational resources and increase students' opportunity costs. Researchers advocate the use of appropriate instructional strategies, improved curriculum design, and remedial lessons in place of grade retention and academic redshirting to improve student learning (Braymen & Piersel, 1987; Cameron & Wilson, 1990; Hauck & Finch(Jr), 1993; Jimerson, 2001; Kundert, et al., 1995). Similar options need to be explored in Bhutan.

6.4 Students' SES and Mathematics Achievement

The relation between the SES index and the Mathematics test scores was analysed by dividing the SES index into quartiles and computing the mean performance scores for each quartile. Table 6.1 shows the result of the analysis.

Table 6. 1. SES Quartiles and Students' Mean Mathematics Test Scores

SES Quartiles	SES Index		Mathematics Test	
	Mean	SE	Mean	SE
1	-1.05	0.02	345.75	5.99
2	-0.53	0.01	359.82	4.70
3	0.13	0.01	362.50	5.93
4	1.45	0.04	374.55	5.50

Table 6.1 shows a pattern that depicts an association between students' SES and performance on the Mathematics test. The difference between the mean Mathematics test scores corresponding to the first quartile of the SES and the mean Mathematics test scores corresponding to the fourth quartile of the SES was significantly different from zero, *diff*=28.80, 95% CI [12.86, 44.73]. The result indicates that the students' SES played an important role in their performance on the Mathematics test.

A linear regression analysis was performed with the students' Mathematics test scores as the dependent variable and SES as the independent variable. A SPSS macro (OECD, 2005c, pp. 214-215) was used for the analysis.

Table 6. 2. Regression of the Mathematics Test Scores on Students' SES

	Statistic	SE
Intercept	360.67	3.86
R-Square	0.03	0.01
SES	11.75	2.49

As shown in Table 6.2, the regression coefficient of SES of 11.75 was significantly different from zero, 95% CI [6.86, 16.63]. A unit change in the index of SES led to an increase of 11.82 on the Mathematics test scores. The SES alone explained 3.0% of the variance in students' performance on the Mathematics test, which was statistically significant, 95% CI [0.38, 4.76].

The analysis of difference in the SES index between female and male students showed no statistically significant difference, *diff* =0.02, 95% CI [-0.11, 0.14], indicating that both boys and girls had similar SES profiles.

Discussion

High-SES students outperformed low-SES students in the Mathematics test. Similar findings are reported in the literature (Caldas & Bankston(III), 1997; Lee, et al., 2007; Mullis, et al., 2004; OECD, 2004a, 2007a; Sirin, 2005; White, 1982). As reported in the OECD (2004a), the performance gap between high-SES students and low-SES students is indicative of inequitable access to educational opportunities. Researchers recommend using educational resources to address the influence of students' SES on achievement (Sirin, 2005; Thrupp & Lupton, 2006), and highlight the schools' potential to bridge the achievement gap between the low- and the high-SES students (Downey, et al., 2004; Sirin, 2005; White, 1982). Policy makers in the Bhutanese education system could use students' SES as one of the criteria for the distribution of educational resources to schools.

6.5 Students' Motivation and Mathematics Achievement

The following analyses were performed on the scale indices of students' motivation: descriptive analysis, analysis by quartiles, correlational analysis, and linear regression analysis. Table 6.3 shows the percentages of students who agreed or strongly agreed with the measures of intrinsic motivation (INMAT) and extrinsic motivation (INSMOT), respectively.

Table 6. 3. Percentages of Students Agreeing or Strongly Agreeing with the Measures of INSMOT and INTMAT

Measures	INSMOT		INTMAT	
	%	SE	%	SE
Making an effort in mathematics is worth it because it will help me in the work that I want to do later on	88.2	2.44		
Learning mathematics is worthwhile for me because it will improve my career prospects or chances	89.4	1.94		
Mathematics is an important subject for me because I need it for what I want to study later on	89.8	2.27		
I will learn many things in mathematics that will help me get a job	91.1	2.38		
I enjoy reading about mathematics			76.5	2.46
I look forward to my mathematics lessons			90.3	2.27
I do mathematics because I enjoy it			83.1	2.60
I am interested in the things I learn in mathematics			89.1	2.39

Overall, students were motivated to learn mathematics as indicated by the high percentages of the students who agreed or strongly agreed with the measures shown in Table 6.3.

Table 6.4 shows the results from the univariate analyses of INSMOT and INTMAT by their quartiles. The difference between the mean Mathematics test scores of the students in the bottom and the top quartiles of INSMOT was statistically significant, *diff*=32.75, 95% CI [17.27, 48.23]. Similarly, the difference in the mean Mathematics test scores of the students in the bottom and the top quartiles of INTMAT was statistically significant, *diff*=41.41, 95% CI [25.16, 57.66].

Table 6. 4. Quartiles of INSMOT and INTMAT with Mean Mathematics Test Scores

Quartiles	INSMOT		Mathematics Test		INTMAT		Mathematics Test	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1	-1.28	.08	343.03	5.72	-1.29	.08	341.91	6.08
2	-.39	.01	356.63	6.63	-.33	.01	356.67	5.64
3	.33	.01	367.06	5.48	.26	.01	360.59	5.83
4	1.25	.02	375.78	5.45	1.25	.02	383.32	5.64

Overall, the univariate analyses of quartiles and performance scores showed that INSMOT and INTMAT had a strong influence on students' performance on the Mathematics test.

A linear regression analysis with INSMOT as the independent variable and the students' Mathematics test scores as the dependent variable showed that the amount of variance explained by INSMOT in students' Mathematics test scores was significantly different from zero, R^2=0.031, 95% CI [0.01, 0.05]. INSMOT explained 3.10% of the variance in the students' Mathematics test scores, with one point gain in INSMOT resulting in an increase of 12.53 on the Mathematics test scores. Similarly, INTMAT also showed a statistically significant R^2=0.047, 95% CI [0.02, 0.08]. INTMAT explained 4.7% of the variance in the students' Mathematics test scores, with a point increase in INTMAT resulting in an increase of 15.48 on the Mathematics test scores. The findings showed that the students' motivation to learn mathematics was strongly related to their performance on the Mathematics test, with strong motivation resulting in higher average test scores.

The Pearson correlation coefficient between INSMOT and INTMAT was statistically significant, *r*=0.60, 95% CI [0.55, 0.65]; showing that INSMOT and INTMAT were complementary. However, the difference between INSMOT and INTMAT was statistically significant with INSMOT surpassing INTMAT, *diff*=1.87, 95% CI [1.68, 2.06].

Analyses of the differences in INSMOT and in INTMAT between female and male students to learn mathematics showed no statistically significant difference, INSMOT, *diff*=0.00, 95% CI [-0.14, 0.14]; INTMAT, *diff*=-0.07, 95% CI [-0.24, 0.09], respectively; indicating that motivation to learn mathematics was similar in female and male students.

Discussion

Analyses of data on motivation showed that, overall, students were high on both intrinsic and extrinsic motivation, and that intrinsic and extrinsic motivation were complimentary. In addition, the analyses showed that motivation related to student performance on the Mathematics test, with students on the high end of the motivation index outperforming students on the low end of the index. While findings of the study on students' motivation are positive, the students who were less motivated needed to be supported with relevant educational interventions. Researchers relate motivation to achievement goal orientation (Schunk, 1996; Vansteenkiste, et al., 2008), self-efficacy (Bandura & Schunk, 1981; Schunk, 1991), self-regulated learning strategies (Ames, 1992; Bell & Kozlowski, 2002; Hidi & Harackiewickz, 2000), and self-concept (Bong & Clark, 1999), indicating these as possible ways to improve motivation in students. Such alternatives need to be pursued by stakeholders in the Bhutanese education system.

Furthermore, the finding that the students' extrinsic motivation to learn mathematics was higher than their intrinsic motivation suggests the lack of persistence in, and conceptual learning of, mathematics (Schunk, 1996; Vansteenkiste, et al., 2008). This implies the need for interventions to promote intrinsic motivation to learn mathematics among the students.

6.6 Students' Self-beliefs and Mathematics Achievement

Students' self-beliefs have been conceptualised in terms of their self-efficacy (MATHEFF) and self-concept (SCMAT) as discussed in Chapter 3. In addition, students' anxiety (ANXMAT) about school subject (e.g., mathematics) has been discussed in the same chapter. Because these three constructs are closely related, they are analysed together in this section. The relationship between these constructs and the student performance on the Mathematics test was analysed by using quartiles, percentage, and regression. Table 6.5 shows the percentages of students who agreed or strongly agreed with the statements on SCMAT and ANXMAT.

Table 6. 5. Percentages of Students Agreeing or Disagreeing with the Measures of SCMAT and ANXMAT

Statements	SCMAT		ANXMAT	
	%	SE	%	SE
I am just not good at mathematics	51.6	1.96		
I get good marks in mathematics	51.4	1.94		
I learn mathematics quickly	43.7	1.60		
I have always believed that mathematics is one of my best subjects	57.0	1.88		
In my mathematics class, I understand even the most difficult work	41.5	1.85		
I often worry that it would be difficult for me in mathematics class			60.2	1.94
I get very tense when I have to do mathematics homework			42.3	1.58
I get very nervous doing mathematics problems			54.4	1.98
I feel helpless when doing a mathematics problem			40.0	1.75
I worry that I will get poor marks in mathematics			58.2	2.04

Table 6.5 indicates that a little more than 50% of students agreed or strongly agreed with most of the statements on SCMAT, indicating that almost 50% of the students had low self-concept in mathematics. Similarly, the percentages of students who agreed or strongly agreed with the statements on ANXMAT were slightly below 50%, indicating that over half of the students were anxious about mathematics. Table 6.6 presents the percentages of students who reported as confident or very confident (high self-efficacy) in doing various mathematical tasks.

Table 6. 6. Percentages of Students Reporting Confident or Very Confident (high self-efficacy) in Solving Mathematical Problems

Mathematical Problems	MATHEFF	
	%	SE
Using a bus timetable to work out how long it would take to get from one place to another	47.0	1.84
Calculating how much cheaper a TV would be after a 30% discount	65.8	2.49
Calculating how many square metres of tiles you need to cover a floor	55.1	1.93
Understanding graphs presented in newspapers	51.0	1.72
Solving an equation like 3x+5=17	79.3	2.26
Finding the actual distance between two places on a map with a 1:10,000 scale	67.5	2.56
Solving an equation like 2(x+3)=(x+3) (x-3)	68.0	2.62
Calculating the petrol consumption rate of a car	42.2	1.93

Overall, Table 6.6 indicates that more than 50% of students reported high MATHEFF at performing various mathematics tasks (e.g., calculating distance, discount, area).

Table 6.7 presents the quartiles of MATHEFF, SCMAT, and ANXMAT with the corresponding mean Mathematics test scores of the students in each quartile.

Table 6. 7. Quartiles of MATHEFF, SCMAT, and ANXMAT with Students' Mean Mathematics Test Scores

Q	MATHEFF		Math Test		SCMAT		Math Test		ANXMAT		Math Test	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1	-1.05	0.02	333.38	6.59	-1.60	0.05	337.15	6.08	-0.99	0.01	392.41	5.97

2	-0.46	0.03	352.37	5.40	-0.40	0.02	348.56	4.92	-0.50	0.01	364.20	5.49
3	0.14	0.01	366.32	5.13	0.42	0.01	364.80	6.20	0.05	0.01	350.10	4.94
4	1.58	0.06	390.39	6.17	1.69	0.05	392.00	5.53	1.33	0.04	335.94	6.15

Note: Q = quartiles

The differences in the mean Mathematics test scores of the students in the bottom and the top quartiles of MATHEFF, SCMAT, and ANXMAT was significantly different from zero; *diff* =57.01, 95% CI [39.20, 74.82], *diff* =54.85, 95% CI [38.73, 70.96], and *diff* =56.47, 95% CI [39.56, 73.39], respectively. The results indicate that MATHEFF, SCMAT, and ANXMAT were related to students' performance on the Mathematics test, with higher self-efficacy and self-concept and lower anxiety resulting in increased average Mathematics test scores.

A linear regression analysis with MATHEFF as the independent variable and students' Mathematics test scores as the dependent variable showed that MATHEFF accounted for 8.9% of the variance in students' performance scores, R^2=0.089, 95% CI [0.05, 0.13]. A unit increase in the score on MATHEFF resulted in a statistically significant increase of 19.60 on the Mathematics test scores, 95% CI [15.34, 23.85]. The result from the regression analysis with SCMAT as the independent variable and students' Mathematics test scores as the dependent variable showed that SCMAT explained 7.8% of the variance in students' Mathematics test scores, R^2= 0.078, 95% CI [0.05, 0.11]. A unit increase in the score on SCMAT led to a statistically significant increase of 15.25 on the Mathematics test scores, 95% CI [12.00, 18.52]. Similarly, the result from the regression analysis with ANXMAT as the independent variable and students' Mathematics test scores as the dependent variable showed that ANXMAT accounted for 6.4% of variance in the Mathematics test scores, R^2=0.064, 95% CI [0.03, 0.10]. A unit decrease in the score on ANXMAT led to a statistically significant increase of 18.91 on the Mathematics test scores, 95% CI [-24.08, -13.75].

The analysis of the difference in the scores of female and male students on MATHEFF showed that the male students had significantly higher scores on MATHEFF than the female students, *diff* =-0.46, 95% [-0.60, -0.32], showing that the male students had a greater sense of self-efficacy with regard to mathematics than the female students. Female students scored significantly higher than male students on ANXMAT, *diff*=0.31, 95% CI [0.19, 0.43], indicating that the female students were more anxious about mathematics than the male students. There was no significant difference in the scores of female and male students on SCMAT, *diff*=0.07, 95% CI [-0.01, 0.16].

Discussion

Results from the analyses of data on students' self-beliefs and anxiety revealed a range of insights.

First, analyses of data on students' self-concept showed that about half of the students had a low self-concept in mathematics, and that students' self-concept related to their performance on the Mathematics test. Similar to the findings reported in Bong and Clark (1999), students with high self-concept outperformed students with low self-concept. Further, Bong and Clark (1999) noted that self-concept is mostly fostered by social comparison through a series of self-evaluations and self-descriptions under varying situations, implying that students develop self-concept based on their experiences in school. This statement provides a valuable reference for designing educational interventions aimed at improving students' self-concept, which may be relevant for Bhutan. Further, students' self-concept may indicate their affective responses to school subject, because the former comprises the latter (Bong & Clark, 1999). This relationship suggests that affective responses of one-half of the Grade 10 Bhutanese students to mathematics were less positive. Because the consequences of negative affects are doubts, anxiety, avoidances, and disengagements (Carver & Scheier, 2005), it is likely that the majority of Grade 10 Bhutanese students will shun mathematics in their further studies.

Second, the results from the analyses of data on self-efficacy revealed that more than one-half of the students had a high sense of self-efficacy in mathematics and that self-efficacy and student performance were related—students with high self-efficacy outperformed students with low self-efficacy. Similar findings are reported in the literature (Bong & Clark, 1999; Hattie, 2009; Pajares, 1996; Schunk, 1991). Students' self-efficacy may be raised by paying attention to its sources. Bandura (1977) stated that self-efficacy is acquired from sources such as performance accomplishments, vicarious experiences, verbal persuasion, and emotional arousal. As suggested by Pajares (1996), these sources of self-efficacy could be used for developing educational interventions to nurture positive self-efficacy in Bhutanese students. Girls would need to be given more attention while developing interventions because they had significantly lower self-efficacy in mathematics than boys.

Third, analyses of data on anxiety showed that half of the students were apprehensive about mathematics. The study showed that anxiety about mathematics was negatively related to student performance on the Mathematics test, supporting the findings reported in the literature (Ashcraft & Moore, 2009; Ma, 1999; Preston, 2008). Given that half the student population was anxious about mathematics, Bhutan may need to develop relevant educational interventions to reduce students' anxiety about mathematics. The interventions may also be made more relevant to girls because they were significantly more anxious of mathematics than boys. Potential references for designing educational interventions for reducing students' anxiety about mathematics are teaching strategies (Alsup, 2004; Furner & Berman, 2003; Hellum-Alexander, 2010; Norwood, 1994), classroom climate (Ma, 1999), and familiarity with symptoms of anxiety (Geist, 2010; Newstead, 1998), in pursuing solutions to the anxiety levels amongst Bhutanese students.

Finally, self-concept, self-efficacy, and anxiety also indicate the level of student engagement (OECD, 2004a, 2007a). The knowledge that about one-half of the students had low self-concept, low self-efficacy, and high anxiety in mathematics suggests that the majority of students were emotionally disengaged in mathematics. This shows that the Bhutanese education system needs to promote student engagement by applying a range of interventions commonly recommended in the literature (Buhs, et al., 2006; Fredricks, et al., 2004; Hughes, et al., 2008; Sullivan, et al., n.d).

6.7 Students' Self-Regulated Learning Strategies and Mathematics Achievement

Memorisation (MEMOR), elaboration (ELAB), and control (CSTRAT) strategies were used as constructs of self-regulated learning strategies as described in Chapter 4. The following analyses were performed on these measures: analysis by quartiles, regression analysis, and gender differences.

The percentages of students who agreed or strongly agreed to various statements that measured MEMOR, ELAB, and CSTRAT are shown in Table 6.8.

Table 6. 8. Percentages of Students Agreeing or Strongly Agreeing with the Measures of MEMOR, ELAB, and CSTRAT

Statements	MEMOR		ELAB		CSTRAT	
	%	SE	%	SE	%	SE
I go over some problems in mathematics so often that I feel as if I could solve them in my sleep	55.9	2.16				
When I study for mathematics, I learn as much as I can off by heart	71.8	1.78				
In order to remember the method for solving a mathematical problem, I go through examples again and again	90.9	2.24				
To learn mathematics, I try to remember every step in a procedure	84.0	2.20				
When I am solving mathematical problems, often think of new ways to get the answer			82.2	2.52		

I think how the mathematics I have learnt can be used in everyday life			87.8	2.47		
I try to understand new concepts in mathematics by relating them to things I already know			84.8	2.43		
When I am solving a mathematical problem, I often think about how the solution might be applied to other interesting questions			82.6	1.90		
When learning mathematics, I try to relate the work to things I have learnt in other subjects			73.8	2.69		
When I study for a mathematics test, I try to work out what are the most important parts to learn					95.2	2.10
When I study mathematics, I make myself check to see if I remember the work I have already done					89.9	1.71
When I study mathematics, I try to figure out which concepts I still have not understood properly					91.1	2.07
When I cannot understand something in mathematics, I always search for more information to clarify the problem					87.2	1.97
When I study mathematics, I start by working out exactly what I need to learn					86.2	1.97

As shown in Table 6.8, all three self-regulated learning strategies appear to be used by the majority of Grade 10 Bhutanese students.

The results from univariate analyses of MEMOR, ELAB, and CSTRAT scale indices by their quartiles are presented in Table 6.9. Students' mean Mathematics test scores for each quartile are also displayed in the same table.

Table 6. 9. Mathematics Test Scores by Quartiles of MEMOR, ELAB, and CSTRAT

Q	MEMOR				ELAB				CSTRAT			
	Score		Math Test		Score		Math Test		Score		Math Test	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
1	-1.25	0.11	359.10	5.55	-1.25	0.11	357.11	6.09	-1.23	0.09	348.21	5.50
2	-0.34	0.01	359.48	6.69	-0.31	0.01	359.78	5.08	-0.41	0.01	362.43	6.36
3	0.21	0.01	361.50	6.19	0.20	0.01	360.82	5.97	0.22	0.01	363.91	5.80
4	1.23	0.03	362.45	5.54	1.25	0.03	364.83	6.52	1.30	0.03	367.97	4.83

Q=quartile; M=mean

The differences in the mean Mathematics test scores of the students in the bottom quartiles and the top quartiles of MEMOR and ELAB were statistically non-significant, *diff*=3.35, 95% CI [-12.02, 18.72], and *diff*=7.72, 95% CI [-9.76, 25.20], respectively. However, the analysis revealed a statistically significant difference in the mean Mathematics test scores of the students in the bottom quartile and the top quartile of CSTRAT, *diff*=19.76, 95% CI [5.41, 34.10].

The result of a linear regression analysis with MEMOR, ELAB, and CSTRAT as the independent variables and the Mathematics test scores as the dependent variable showed that the variance explained by the three variables in students' performance on the Mathematics test was not significantly different from zero, $R^2 = 0.01$, 95% CI [0.00, 0.03]. The regression coefficient of CSTRAT was significantly different from zero, *B*=6.94, 95% CI [2.75, 11.12]. A unit increase in score on CSTRAT led to an increase of 7 on the Mathematics test scores. The regression coefficients of MEMOR and ELAB were negative and were not significantly different from zero, *diff*= -0.85, 95% CI [-4.27, 1.97] and *diff*=3.54, 95% CI [-7.57, 2.36], respectively. A unit decrease in the scores on MEMOR and ELAB, respectively, led to an increase of two and three on the Mathematics test scores.

Discussion

The findings from the study on the students' use of self-regulated learning strategies showed that the majority of Bhutanese students used them to improve their learning. Of the three aspects of self-regulation strategies, only control strategies resulted in a significant difference in the performance of the students on the Mathematics test, with the difference in favour of students who reported using the strategy. The absence of a relation between memorization or elaboration and student performance is known to depend on other moderating factors such as students' self-efficacy and motivation, indicating the need for a holistic approach to enhancing students' capacity to use self-regulated learning strategies (Eccles & Wigfield, 2002; Pintrich & De Groot, 1990; Zimmerman, et al., 1992).

Although, the majority of Bhutanese students reported using self-regulated learning strategies, the presence of a small percentage of Bhutanese students who reported not using them and the lack of statistically significant relationships of memorisation and elaboration strategies to the Mathematics test scores are indicative of the absence of formal educational programmes on self-regulated learning strategies in Bhutanese schools. Because self-regulated learning strategies are widely regarded as necessary pre-conditions for meaningful, effective, and successful life-long learning (Puustinen & Pulkkinen, 2001), the absence of educational programmes on self-regulated learning strategies in Bhutanese schools suggests that Bhutanese students might complete their compulsory education without having acquired most of the attributes of a life-long learner that have been convincingly reported in the literature (Knapper & Cropley, 2000; Resnick, 1987; Secretary of Labour's Commission on Achieving Necessary Skills, 1991; Tuijnman, et al., 1997). This in turn suggests the lack of preparedness of Bhutanese students for future challenges. Therefore, it is imperative that opportunities are provided for all Bhutanese students to learn and apply self-regulated learning strategies.

6.8 Students' Preferences for Learning Environments and Mathematics Achievement

Students' preferences for learning environments were divided into competitive (COMPLRN) or cooperative (COOPLRN) learning environments as in Chapter 4. Percentages, quartiles, and regression analyses were preformed on COMPLRN and COOPLRN to study their associations with the student achievement in the Mathematics test. Table 6.10 presents the percentages of students who agreed or strongly agreed to the measures of COMPLRN and COOPLRN.

Table 6. 10. Percentages of Students Agreeing or Strongly Agreeing with the Measures of COMPLRN and COOPLRN

Measures	COMPLRN		COOPLRN	
	%	SE	%	SE
I would like to be the best in my class in mathematics	82.4	2.31		
I try very hard in mathematics because I want to do better in the exams than the others	93.2	2.24		
I make a real effort in mathematics because I want to be one of the best	87.6	2.18		

In mathematics I always try to do better than other students in my class	65.8	1.95		
I do my best work in mathematics when I try to do better than others	80.2	2.29		
In mathematics I enjoy working with other students in groups			83.1	2.41
When we work on a project in mathematics, I think that it is a good idea to combine the ideas of all the students in a group			92.3	2.27
I do my best work in mathematics when I work with other students			80.6	1.89
In mathematics I enjoy helping others to work well in a group			73.7	2.15
In mathematics I learn most when I work with other students in my class			84.3	2.29

As shown in Table 6.10, both competitive and cooperative learning environments are popular preferences among Grade 10 Bhutanese students.

Table 6.11 shows the quartiles of COMPLRN and COOPLRN with the students' mean Mathematics test scores.

Table 6. 11. Mean Mathematics Test Scores by Quartiles of COMPLRN and COOPLRN

Quartiles	COMPLRN		Mathematics Test		COOPLRN		Mathematics Test	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1	0.29	0.07	350.63	5.92	-1.16	0.04	359.24	5.74
2	1.35	0.01	356.97	6.12	0.04	0.01	359.67	6.6
3	2.29	0.01	362.13	5.25	0.89	0.01	359.87	6.02
4	3.97	0.04	372.78	5.99	2.37	0.05	362.55	5.42

The difference between the mathematics performance scores of the students in the bottom and the top quartiles of COMPLRN was statistically different from zero, *diff*=22.15, 95% CI [5.64, 38.66], indicating that students with preference for COMPLRN performed better on the Mathematics test. The difference between the performance scores of the students in the bottom and the top quartiles of

COOPLRN was statistically not significant, *diff*=3.31, 95% CI [-12.16, 18.79], showing that students with preference for COOPLRN did not perform better on the Mathematics test.

The linear regression analysis with COMPLRN as the independent variable and the students' Mathematics test scores as the dependent variable indicated that the variance explained by COMPLRN in the students' test scores was not statistically significant, $R^2 = 0.013,$ CI [-0.00, 0.03]. Similarly, the result from the regression analysis with COOPLRN as the independent variable and the students' Mathematics test scores as the dependent variable showed that the variance explained by COOPLRN was not statistically significant, $R^2 = 0.00,$ CI [-0.00, 0.01].

The analysis of the difference between the students' mean scores on the COMPLRN and COOPLRN scales resulted in a statistically significant difference, with the difference favouring COMPLRN, *diff*=0.95, 95% CI [0.79, 1.10].

Discussion

The analyses of students' learning preferences resulted in a number of interesting findings. First, the difference between the mean Mathematics test scores of the students in the top and the bottom quartiles of the index of the competitive learning environment was statistically significant, while similar analysis with the index of cooperative learning situations was statistically non-significant. In other words, students who reported a competitive learning preference with their classmates performed better than students who reported a collaborative learning preference with classmates. Second, the difference between the means of the indices of cooperative and the competitive learning preferences was statistically significant, with the competitive learning preference outweighing the cooperative learning preference. These observations have wide-ranging implications for students' development. Stapel and Koomen (2005) noted that students can either adopt a competitive or a cooperative learning preference based on how they view themselves compared with higher- and lower-performing peers, and they claimed that these views form the basis of students' self-evaluation. Self-evaluation from either competitive or cooperative learning preferences has consequences on students' motivational and goal orientations, with the competitive learning preferences related to instrumental motivation and performance goals, and the cooperative learning preferences related to intrinsic motivation and mastery goals (Covington & Omelich, 1984). Because the findings indicate that Grade 10 Bhutanese students preferred competitive learning situations to collaborative learning situations, it may be inferred that most of the students were instrumentally motivated and performance goal-oriented, which may be an indication of disengagement with learning. Because, competitive learning preferences have been linked to structured teaching strategies and cooperative learning preferences to constructivist teaching strategies (Covington & Omelich, 1984; Ediger, 1996), one way to change Bhutanese students' preferences for learning situations could be through frequent use of constructivist teaching strategies by classroom teachers.

6.9 Students' ICT Knowledge and Skills and Mathematics Achievement

Students' ICT knowledge and skills were assessed in terms of the following dimensions: students' experience with ICT; places where students access ICT; duration of students' use of ICT; students' proficiency in performing ICT functions; students' confidence in ICT; and students' attitude towards ICT.

6.9.1 Students' Experience with ICT

Overall, 62.9% (SE=2.87) of the students reported having used computers. Disaggregation by gender found that 59.6% (SE=3.01) of female and 63.0% (SE=3.33) of male students reported having used computers.

The difference in the mean Mathematics test scores of students who reported having used computers and students who reported never having used computers was statistically significant, *diff*=29.61, CI [17.16, 42.06] with the difference in favour of the students who used computers.

Discussion

About four in 10 students reported that they had never used computers. This is not surprising because schools are just beginning to introduce computer studies as a formal subject; therefore, it would take some time before all Grade 10 Bhutanese students had access to computers in schools. However, it is desirable for all Grade 10 students to be able to use computers to perform various ICT functions, because Grade 10 is the end of the basic education in Bhutan. ICT knowledge and skills have become indispensable for competent participation in the world of work and for pursuing life-long learning (OECD, 2005b).

Students who used computers performed better on the Mathematics test than students who did not use computers. Similar findings were reported in Mullis et al. (2008) and the OECD (2004a). In addition, experience with ICT was reported to: increase students' work expediency, improve productivity, foster independence and peer support, broaden the information resource base, and maintain currency of knowledge and information (Ruthven, et al., 2004). The use of ICT in schools was also reported to have positive influences on students' self-regulatory learning strategies (Ilomaki & Rantanen, 2007; Lim, 2002). All these indicate the need for all students to be able to access ICT facilities in Bhutanese schools.

6.9.2 Places where Students Access ICT

Students who had experience of using computers were asked about locale of access. Table 6.12 shows the percentages of students who reported having access to computers at home, school, or other places.

Table 6. 12. Percentages of Students with Access to Computers at Different Locations

Place	%	SE
Home	23.0	2.41
School	81.0	2.41
Other places	53.4	2.98

As shown in Table 6.12, the majority of students who used computers had access to computers at their school rather than home or other places.

Students were also asked how often they used computers at these places. The frequency of computer use was categorised into (a) frequent use, (b) moderate use, and (c) rare or no use. Students were considered to make frequent use of computers if they reported using computers almost every day or a few times each week. The moderate use of computers comprised students' who reported using computers between once a week and once a month. The rare or no use of computers included students who reported the use of computers less than once a month or never. Table 6.13 presents the percentages of students by the frequency of using computers at different locations.

Table 6. 13. Percentages of Students by Frequency of Using Computers at Different Places

Place	Frequent Use		Moderate Use		Rare or No Use	
	%	SE	%	SE	%	SE
Home	29.7	2.69	4.3	0.73	65.9	3.54
School	61.3	3.39	11.0	1.47	27.6	3.48
Other places	25.5	2.40	16.2	1.37	58.3	3.30

Table 6.13 shows that the majority of students (61.3%) who reported using computers frequently used computers at school much more than at home or other places.

Table 6.14 shows the result of a bivariate analysis of the relationship between the frequency with which students used computers at different places and their performance on the Mathematics test.

Table 6. 14. Frequency of Computer Use at Different Locations and Mean Mathematics Test Scores

Location	Frequent Use		Moderate Use		Rare or No Use	
	Mean Score	SE	Mean Score	SE	Mean Score	SE
Home	380.54	6.18	351.38	17.22	376.8	4.98
School	386.92	4.66	361.72	10.38	376.20	8.66
Other places	368.48	7.03	376.82	8.10	381.03	4.97

As shown in Table 6.14, reading horizontally, a frequent use of computers at home and school was associated with marginally higher mean Mathematics test score. The use of computers by students at other places revealed a negative association with their performance on the Mathematics test, with the students reporting rare or no use of computers scoring higher than the students reporting moderate or frequent use of computers.

An interesting insight appears when Table 6.14 is read vertically. While numbers remain small, it is noteworthy that students who made rare or no use of computers outperformed students who made moderate use of computers at whatever site. In addition, students who made frequent use of computers at home were outperformed by students who made rare or no use of computers in other places. There may well be implications in this for some further investigation by stakeholders, particularly with regard to training to maximise the use of ICT facilities.

Discussion

The use of computers at schools by students outstripped their use of computers at other places. This indicates the important role of schools as providers of equitable access to ICT learning opportunities because most students were either not able to afford home computers or access computers at other places. This is a compelling call for the Ministry of Education to invest more heavily in ICT-related school resources, both in training and in provision of facilities.

The students' frequent use of ICT at other places (e.g., Internet cafes, friends' places) was negatively associated with their performance on the Mathematics test. This observation may be attributable to less monitoring of, and support for, the use of ICT available to students in such places compared to the monitoring and support by teachers and parents that the students often receive when using ICT at their homes and schools.

6.9.3 Duration of Students' Use of ICT

Students were asked how long they have been using computers. Table 6.15 shows the percentages of students who reported using computers for different lengths of time.

Table 6. 15. Percentages of Students and Length of Time They have been Using Computers

Length of Time for using Computers	%	SE
Less than one year	48.1	2.99
One to three years	38.5	2.80
Three to five years	7.1	1.15
More than five years	6.4	1.11

Table 6.15 indicates that the majority of the Grade 10 students have been using computers for three or fewer than two years.

Table 6.16 shows the result of the bivariate analysis of the relationship between the length of time students reported using computers and their Mathematics test scores.

Table 6. 16. Duration of Using ICT and Students' Mathematics Test Scores

Length of Time having used Computers	Mean Score	SE
Less than one year	363.25	5.21
One to three years	378.79	6.41
Three to five years	394.55	10.97
More than five years	383.49	12.80

Table 6.16 indicates that the length of students' use of ICT is positively related to their performance on the Mathematics test.

Discussion

The analysis of the duration of the use of ICT by students revealed that the majority of students began using ICT towards the end of their basic education. This may be indicative of schools' preferential policy on the use of ICT by their students. For instance, schools might have offered comparatively less access to ICT to students in lower grades, compared to Grade 10 students, because of limited ICT facilities. However, the finding that the duration of students' use of ICT was directly related to student achievement implies that schools could consider extending, wherever possible, access to ICT to students in lower grades. A prolonged use of ICT in Bhutanese schools may translate to improvements in student achievement in school subjects.

6.9.4 Students' Proficiency in Performing ICT Functions

ICT functions were grouped into two major groups: ICT for the Internet and entertainment (INTUSE) and ICT for computer programs and software (PRGUSE). Students were asked how frequently they used ICT for INTUSE and PRGUSE. Table 6.17 shows the percentages of students who reported making frequent use of ICT for INTUSE and PRGUSE.

Table 6. 17. Percentages of Students Reporting Frequent Use of ICT for INTUSE PRGUSE

INTUSE	%	SE
The Internet to look up information	38.3	2.48
Games on a computer?	48.7	2.57
The Internet to collaborate with a group or team?	23.9	2.10
The Internet to download software (including games)?	32.0	2.12
The Internet to download music?	39.0	2.80
A computer for electronic communication (e.g., e-mail or chat rooms)?	29.8	2.40
PRGUSE		
Word processing (e.g., MS Word)?	51.1	3.07
Spreadsheets (e.g., MS Excel)?	37.3	2.50
Drawing, painting, or graphic programs on a computer?	52.0	2.93
Educational software such as mathematics programs?	28.5	1.92
The computer to help you learn school material?	42.4	2.45
The computer for programming?	40.3	2.22

Overall, Table 6.17 shows that, except for word processing and graphic activities, fewer than 50% of students made frequent use of ICT for INTUSE or PRGUSE.

Analyses of associations of INTUSE and PRGUSE with students' performance on the Mathematics test showed interesting results. The difference in the mean Mathematics test scores of the students in the top and the bottom quartiles of INTUSE was statistically significant in favour of the students in the top quartile, *diff*=25.36, 95% CI [8.50, 42.21]. Similarly, the difference in the mean Mathematics test scores of the students in the top and the bottom quartiles of PRGUSE was statistically significant in favour of the students in the top quartiles, *diff*=32.41, 95% CI [14.61, 50.20].

Further, the gender difference in INTUSE was statistically significant in favour of male students, *diff* =-0.13, 95% CI [-0.24, -0.01]. The gender difference in PRGUSE was statistically non-significant, *diff*=0.06, 95% CI [-0.09, 0.21].

Discussion

The proportion of students who used ICT for the Internet and entertainment was almost equal to the number of students who did not use ICT for these reasons. Similarly, the number of students who used ICT for computer programs or software was almost equal to the number of students who did not use ICT for computer programs or software. These observations indicate the lack of ICT in the classroom, the lack of teaching and learning activities that require students to depend on ICT, and the lack of uniform access to ICT in schools. Students who reported rarely or not having used ICT were at risk of being marginalised academically because their performance on the Mathematics test was lower than the performance of the students who reported using ICT more frequently.

The finding that boys used the Internet more frequently than girls is an interesting insight that signals the need for further studies into the subject. In the context of Bhutan, this finding means that boys had availed themselves of access to the Internet more often than girls. However, attitudes towards ICT, confidence in ICT, anxiety about ICT, and socio-cultural environments are other possible factors for the difference in the use of the Internet by boys and girls.

6.9.5 Students' Confidence in ICT

Responses to the question of confidence in computer use was collected from students by asking them how well they could do routine ICT tasks (e.g., turning on a computer), Internet tasks (e.g., getting on to the Internet), and high-level tasks (e.g, getting rid of computer viruses).

Students were considered highly confident if they reported "I can do this very well by myself", moderately confident if they reported "I can do this with help from someone", and less confident if they reported "I know what this means, but I cannot do it" or "I do not know what this means". Table 6.18 presents the percentages of students reporting how well they could perform routine tasks, Internet tasks, and high-level tasks.

Table 6. 18. Percentages of Students Reporting How Well They can Perform Routine, Internet, and High-Level Tasks

Routine Tasks	1*		2*		3*		4*	
	%	SE	%	SE	%	SE	%	SE
Start a computer	70.0	2.14	19.3	1.76	6.7	1.20	4.0	0.88
Open a file	76.3	2.03	16.0	1.53	4.2	0.80	3.5	0.66
Create/edit document	64.9	2.56	21.9	1.96	6.4	1.04	6.8	1.30
Scroll a document up and down a screen	53.9	2.55	25.0	1.99	9.8	1.14	13.5	1.99
Copy a file from a floppy disk	46.4	2.84	28.8	1.77	13.7	1.27	11.1	1.37
Save a computer document or file	67.4	2.52	19.0	1.78	6.5	0.83	7.0	1.44
Print a computer document or file	52.9	2.63	28.2	1.95	11.5	1.37	7.4	1.64
Delete a computer document or file	70.6	2.28	17.7	1.74	6.9	1.08	4.7	0.82
Move files from one place to another on a computer	57.8	2.52	23.3	2.08	10.9	1.43	8.1	1.42
Play computer games	68.8	2.19	19.1	1.58	5.7	0.71	6.4	1.17
Draw pictures using mouse	69.5	1.53	21.6	1.38	4.0	0.63	5.0	0.89
Internet Tasks								
Get on to the Internet	43.7	3.45	29.1	2.15	13.9	1.44	10.3	1.51
Copy or download files from the Internet	31.7	2.90	36.1	2.56	17.3	1.95	14.9	1.88
Attach a file to an e-mail message	25.1	2.39	38.5	2.36	17.1	1.90	19.2	2.09
Download music from the Internet	38.0	2.56	35.2	2.00	13.0	1.51	13.8	1.97
Write and send e-mails	38.4	3.21	32.7	1.91	15.6	1.42	13.3	1.74
High-Level Tasks								
Use software to find and get rid of computer viruses	21.8	2.12	44.1	2.02	18.6	1.47	15.5	1.99
Use a database to produce a list of addresses	45.0	3.12	28.3	2.12	13.3	1.41	13.5	1.99

Create a computer program	32.3	2.18	36.3	2.21	14.2	1.47	17.2	2.10
Use a spreadsheet to plot a graph	31.3	2.00	32.2	2.05	16.0	1.25	20.5	2.08
Create a presentation (e.g., using PowerPoint)	51.3	2.67	23.0	2.14	10.9	1.31	14.8	1.87
Create a multi-media presentation (with sound, pictures, video)	34.4	2.46	34.6	1.92	15.4	1.26	15.6	1.62
Construct a web page	16.3	1.70	37.1	1.73	23.1	1.67	23.5	2.25

Note: 1*= I can do this very well by myself; 2*= I can do this with help from someone;
3*= I know what this means, but I cannot do it; and 4=* I do not know what this means

Table 6.18 shows that at least 46.4% of students who reported using computers had high levels of confidence in performing routine ICT tasks. At least 25.1% of students reported having high confidence at each of the five Internet tasks. More than 16.3% of students reported having high confidence at performing high-level tasks.

Table 6.19 shows the result of the bivariate analyses of the quartiles of the indices of students' confidence in performing routine (ROUTCONF), Internet (INTCONF), and high-level (HIGHCONF) tasks and students' performance on the Mathematics test.

Table 6. 19. Students' Mean Mathematics Test Scores by Quartiles of Confidence in ICT Functions

Q	Routine Tasks				Internet Tasks				High-Level Tasks			
	Mean Index	SE	Mean Score	SE	Mean Index	SE	Mean Score	SE	Mean Index	SE	Mean Score	SE
1	-2.05	.05	355.64	6.94	-2.78	.07	372.47	7.68	-5.91	.00	342.77	6.13
2	-0.86	.02	365.18	6.61	-.91	.03	365.30	7.54	-4.70	.14	353.70	5.41
3	0.24	.03	379.28	8.06	.18	.03	374.37	7.26	-.02	.03	372.92	6.03
4	1.73	.03	394.33	7.34	2.38	.06	381.96	7.09	1.80	.07	373.10	5.80

Q=quartile

As shown in Table 6.19, there is a clear pattern in the mean Mathematics test scores of the students in the top and the bottom quartiles of the students' confidence in routine tasks, with the students in the top quartile scoring significantly higher than the students in the bottom quartile, *diff*=38.90, 95% CI [18.88, 58.49]. Similarly, the mathematics performance scores of the students in the top and the bottom quartiles of the index of high-level tasks was significantly different, *diff*=30.33, 95% CI [13.68, 46.98]. There was no equivalent significance for Internet tasks.

Discussion

Overall, students were relatively more confident at performing routine ICT tasks on computers than using the Internet or performing high-level tasks. This pattern is consistent with the pattern reported by the OECD (2005a). Across the OECD countries that participated in PISA 2003; on average, at least 65%, 58%, and 21% of students had high confidence at performing routine, Internet, and high-level tasks, respectively, on the computer. However, the percentages of students having high confidence across the OECD countries in performing routine, Internet, and high-level tasks on computer are greater than the corresponding percentages of students in Bhutan.

Students' confidence in routine and high-level tasks was associated with a significant performance difference on the Mathematics test, with the difference in favour of confident students. This finding indicates that improved ICT provision may contribute positively to student learning in mathematics in Bhutan.

6.9.6 Students' Attitude Towards ICT

To measure their attitudes towards ICT, students were asked to rate their agreement with four response statements on a four-category Likert-type item, with the categories ranging from 1 (strongly agree) to 4 (strongly disagree). Table 6.20 presents the percentages of students who responded agree or strongly agree with the four response statements.

Table 6. 20. Percentages of Students Agreeing or Strongly Agreeing with Attitude Statements

Statement	%	SE
It is very important to me to work with a computer	94.9	2.50
I think playing with or working with a computer is really fun	77.0	2.99
I use a computer because I am very interested	95.1	2.77
I lose track of time when I am working with the computer	66.3	2.79

As shown in Table 6.20, more than 66.3% of students had favourable attitudes towards ICT.

Gender difference in the index of students' attitude towards ICT (ATICOMP) was statistically not significant, *diff*=-0.05, 95% CI [-0.24, 0.14], indicating that both male and female students had similarly favourable attitudes towards ICT. In addition, a bivariate analysis of the relation between

ATICOMP and students' performance on the Mathematics test by using quartiles of ATICOMP showed that the difference in the performance of the students in the top and the bottom quartiles of the index was statistically non-significant, *diff*=13.84, 95% CI [-4.51, 32.16].

Discussion

The majority of students, about two-thirds of the sample population, reported positive attitudes towards ICT. This finding indicates that students, irrespective of their gender, have high interest in using ICT for performing various ICT functions.

6.10 Students' Views of Classroom Management and Mathematics Achievement

Teacher support (TEACHSUP) and classroom disciplinary climate (DISCLIM) were used as factors of classroom management. The following analyses were performed on TEACHSUP and DISCLIM: descriptive, quartiles, regression, and correlation.

Table 6.21 shows the percentages of students who reported observing the measures of TEACHSUP and DISCLIM in every lesson or in most of the lessons.

Table 6. 21. Percentages of Students Observing the following Measures in Every Lesson or in Most of the Lessons

Measures	TEACHSUP		DISCLIM	
	%	SE	%	SE
The teacher shows an interest in every student's learning	78.70	4.24		
The teacher gives extra help when students need it	86.59	2.20		
The teacher helps students with their learning	84.35	2.00		
The teacher continues teaching until the students understand	81.53	2.23		
The teacher gives students an opportunity to express opinions	70.20	2.22		
Students don't listen to what the teacher says			14.10	1.18
There is noise and disorder in the classroom			17.86	1.47
The teacher has to wait a long time for students to quieten down			26.55	1.50
Students cannot work well			17.01	1.51
Students don't start working for a long time after the lesson begins			24.80	1.38

As shown in Table 6.21, students reported a high observation of TEACHSUP and a low observation of DISCLIM.

Results from the univariate analyses of students' performance by the quartiles of TEACHSUP and DISCLIM, respectively, are shown in Table 6.22.

Table 6. 22. Students' Mean Mathematics Test Scores by Quartiles of TEACHSUP and DISCLIM

Quartiles	TEACHSUP		Mathematics Test		DISCLIM		Mathematics Test	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1	-1.19	0.10	354.79	5.26	-1.19	0.12	352.48	5.40
2	-0.32	0.01	357.95	7.09	-0.26	0.01	359.87	6.01
3	0.19	0.01	361.15	5.93	0.22	0.01	359.89	5.78
4	1.30	0.02	368.67	6.22	1.19	0.04	370.30	6.32

The difference in the mean Mathematics test scores of students in the bottom and the top quartiles of TEACHSUP was statistically not significant, *diff*=13.90, 95% CI [-2.21, 29.98]. Conversely, the difference in the mean Mathematics test scores of students in the bottom and the top quartiles of DISCLIM was significantly different from zero; *diff*=17.82, 95% CI [1.53, 34.11], indicating that the positive classroom disciplinary climate is related to student learning.

Results from a linear regression analysis of TEACHSUP as an independent variable and students' performance on the Mathematics test as a dependent variable showed that the variance explained by TEACHSUP was not significantly different from zero, $R^2 = 0.00$, 95% CI [0.00, 0.00]. Similarly, the regression analysis of DISCLIM as an independent variable and students' performance on the Mathematics test as a dependent variable revealed that the amount of variance explained by DISCLIM in the Mathematics test scores was statistically not significant, $R^2 = 0.01$, 95% CI [-0.01, 0.02].

The analysis of difference in the scores of female and male students on TEACHSUP showed that both female and male students reported similar observations of TEACHSUP, *diff* =0.02, 95% CI [-0.13, 0.17,]. The difference in the scores of female and male students on DISCLIM was significantly different from zero, *diff* =0.17, 95% CI [0.01, 0.32,], indicating that the female students reported more favourable experience with classroom disciplinary matters.

Discussion

Analyses of data on classroom management, expressed in terms of teacher support and classroom disciplinary climate, showed that the majority of students perceived classroom management to be student friendly. Students reported frequent teacher support and fewer disciplinary problems in the classroom. This is encouraging because effective classroom management facilities successful teaching and learning activities in the classroom and greater student behavioural engagement (Cothran, et al., 2003; Marzano, 2003b). However, teacher support failed to relate to student performance on the Mathematics test. The finding indicates the likelihood of some students having falsely identified with positive statements about their teachers. On the other hand, as expected, the students who reported less frequent classroom disciplinary problems outperformed the students who reported more frequent classroom disciplinary problems. The influence of classroom disciplinary problems on student learning indicates the need for teachers to practice their classroom management knowledge and skills. The literature offers a range of evidence-based interventions on classroom management for contemplation (Akin-Little, et al., 2007; Simonsen, et al., 2008; Zuckerman, 2007). Specifically, setting classroom rules, developing disciplinary interventions, building supportive teacher-student relationships, and an appropriate mental set are frequently linked to efficient classroom management (Marzano, 2003a, 2003b), and they could be productively applied more often in Bhutanese schools.

6.11 Students' Views of School Climate and Mathematics Achievement

School climate was measured by using the students' attitude towards their school (ATSCHL), students' sense of belongingness to their school (BELONG), and student and teacher relationship (STUREL). The validity and reliability of these variables were discussed in Chapter 4. Analyses by quartiles, regression, and correlation were used to study the relation between school climate and mathematics achievement.

Table 6.23 shows the percentages of students who agreed or strongly agreed with the statements that measured ATSCHL, BELONG, and STUREL.

Table 6. 23. Percentages of Students Agreeing or Strongly Agreeing to the Measures of ATSCHL, BELONG, and STRUEL

Statements	ATSCHL		BELONG		STUREL	
	%	SE	%	SE	%	SE
School has done little to prepare me for adult life when I leave school	13.3	1.17				
School has been a waste of time	3.3	0.56				

School has helped give me confidence to make decisions	97.7	2.18				
School has taught me things which could be useful in a job	96.6	1.88				
I feel like an outsider (or left out of things).			23.1	1.64		
I make friends easily			87.0	1.77		
I feel like I belong			84.9	2.44		
I feel awkward and out of place			31.0	1.59		
Other students seem to like me			75.2	2.20		
I feel lonely			20.2	1.30		
Students get along well with most teachers					80.6	2.07
Most teachers are interested in students' well-being					90.7	2.45
Most of my teachers really listen to what I have to say					76.4	1.80
If I need extra help, I will receive it from my teachers					95.9	2.59
Most of my teachers treat me fairly					87.7	2.00

Results from the univariate analyses of students' performance by the quartiles of ATSCHL, BELONG, and STUREL are shown in Table 6.24. The differences in mean Mathematics test scores of students in the bottom and the top quartiles of ATSCHL and STUREL were not significantly different from zero, *diff*=11.07, 95% CI [-5.70, 24.84] and *diff*=3.73, 95% CI [-13.74, 21.20], respectively. However, the difference in the mean Mathematics test score of students in the bottom and the top quartiles of BELONG was statistically significant, *diff*=37.12, 95% CI [21.21, 53.03].

Table 6. 24. Students' Mean Mathematics Test Scores by Quartiles of ATSCHL, BELONG, and STRUEL

Q	ATSCHL				BELONG				STUREL			
	Score		Math Test		Score		Math Test		Score		Math Test	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1	0.04	0.02	355.93	5.31	-1.12	0.10	340.96	5.73	0.00	0.02	358.75	6.53
2	0.74	0.01	361.01	6.15	-0.32	0.01	355.93	5.83	0.86	0.02	359.33	8.48
3	1.15	0.01	360.27	6.52	0.14	0.01	367.54	6.17	1.62	0.02	361.82	6.49
4	2.12	0.03	367.00	6.71	1.23	0.04	378.08	5.75	3.17	0.05	362.48	6.06

Q=quartile

The regression analysis with ATSCHL, BELONG, and STUREL as independent variables and the Mathematics test scores as dependent variable showed that ATSCHL, BELONG, and STUREL together explained 4.34% of variance in the students' Mathematics test scores, which was significantly different from zero, $R^2 = 0.04$, 95% CI [0.07, 0.02]. The regression coefficients of ATSCHL and STUREL were positive, but not statistically significant, *B*=4.74, 95% CI [-3.69, 10.10] and *B*=3.63, 95% CI [-3.51, 3.66], respectively. The regression coefficient of BELONG was statistically significant, *B*=16.00, 95% CI [10.34, 21.63]. A unit increase in the score on BELONG corresponds to an increase of 16 on the Mathematics test scores.

The analysis of the difference in the scores of female and male students on ATSCHL showed that the female students reported a statistically significant more favourable attitude towards their school compared with the male students, d*iff*=0.19, 95% CI [0.05, 0.34]. Both female and male students reported similar views about BELONG and STUREL.

Discussion

Overall, the findings from the analyses of data on school climate indicated that the majority of students had positive attitudes towards, and a strong sense of "belonging" to, their schools. In addition, the majority of students had good relationships with their teachers. Because these aspects of school climate indicate student behavioural engagement, it may be inferred that the majority of Bhutanese Grade 10 students were behaviourally engaged in school (Finn & Voelkl, 1993; Fredricks, et al., 2004). A positive school climate also indicate positive affective responses of students to schooling, implying that the majority of Grade 10 Bhutanese students were appreciative of their schooling as they complete their basic education.

Furthermore, the students who identified more positively with the measures of belonging outperformed on the Mathematics test the students who identified less positively. Similar observations have been reported in other studies (C. S. Anderson, 1982; Brand, et al., 2003; Hoy & Hannum, 1997; Macneil, et al., 2007; Marzano, 2001; Sammons, et al., 1995; Scheerens, 2000). Given that school climate is related to student achievement, research-based educational interventions need to be introduced in Bhutanese schools to improve the perspectives of those students who felt schooling was a waste of time and that being in school made them feel awkward and out of place.

6.12 Students' Homework and Mathematics Achievement

Students reported on the frequency with which their teachers assigned mathematics homework and the amount of time that they spent on the homework. These reports were used as indicators of the homework frequency and the time spent on homework. Table 6.25 presents the percentages of students who were assigned mathematics homework with different frequencies by their teachers.

Table 6. 25. Percentage of Students by Homework Frequency and Mean Mathematics Test Scores

Homework			Mathematics Test Scores	
Frequency	%	SE	Mean	SE
Every day	50.78	3.96	360.87	5.60
3 or 4 times a week	36.67	3.00	361.59	5.08
1 or 2 times a week	10.31	1.76	355.21	9.66
Less than once a week	1.84	0.57	363.88	10.83
Never	0.35	0.18	375.66	26.14

As indicated in Table 6.25, fewer than 3% of the students were assigned homework less than once a week or never, which was highly unlikely to be true. The standard errors of the mean Mathematics test scores of these students are also comparatively higher than those of other students who responded in other categories as depicted in Table 6.25, indicating the likelihood of low reliability. On the other hand, over 97% of the students reported that they were given homework in mathematics every day, or three or four times a week, or one or two times a week, with the mean Mathematics test scores of students who reported being assigned homework three to four times a week tending to peak.

While assigning homework by teachers is a good practice, the time spent on homework by students is another consideration. Table 6.26 shows the percentages of students who reported spending different amounts of time on homework.

Table 6. 26. Percentages of Students by Homework Time and Mean Mathematics Test Scores

Homework			Mathematics Test Score	
Time	Percent	SE	Mean	SE
Fewer than 15 minutes	7.66	0.96	352.13	9.21
15-30 minutes	31.56	3.84	362.86	5.93
31-60 minutes	30.82	2.38	366.7	5.13
61-90 minutes	11.57	1.61	361.20	7.89
More than 90 minutes	18.35	2.33	350.92	6.22

Table 6.26 shows that almost two-thirds of students spent between 15-60 minutes on homework tasks, and the performance scores of these students show a marked peak.

The combined result from Tables 6.25 and 6.26 indicates the most effective frequency of homework to be in the range of three to four times a week, and the most effective time required for completing any homework tasks to be in the range of 15 to 60 minutes.

Besides homework, students were asked to report on the amount of time that they had spent per week on doing various activities related to mathematics. Table 6.27 shows the percentages of students who reported doing various activities related to the learning of mathematics for more than two hours per week.

Table 6. 27. Percentages of Students Spending more than Two Hours on Activities Related to Learning Mathematics

Activities	%	SE
Homework or other study set by your mathematics teacher	63.9	2.41
Remedial classes in mathematics at school	21.9	1.39
Enrichment classes in mathematics at school	17.1	1.27
Work with a mathematics tutor	21.8	1.22
Attending out-of-school mathematics classes	15.8	0.78
Other mathematics activities (e.g., mathematics competitions, mathematics club)	11.2	1.07

As indicated in Table 6.27, except for homework, not many students spent more than two hours per week on the activities related to learning mathematics. On the other hand, the percentages of students who reported attending different out-of-school-time lessons were high, as indicated in Table 6.28.

Table 6. 28. Percentages of Students Attending the following Out-of-School-Time Lessons

Activities	%	SE
One to one lessons with a teacher who is also a teacher at your school	54.4	2.07
One to one lessons with a teacher who is not a teacher at your school	25.9	1.69
Lessons in small groups (less than 8 students) with a teacher who is also a teacher at your school	49.0	2.42
Lessons in small groups (less than 8 students) with a teacher who is not a teacher at your school	19.1	1.35
Lessons in larger groups (8 students or more) with a teacher who is also a teacher at your school	52.4	2.60
Lessons in larger groups (8 students or more) with a teacher who is not a teacher at your school	16.2	1.18

Table 6.27 and Table 6.28 show that the majority of students did not get extended learning opportunities during school time, while many students did engage in taking out-of-school-time lessons. This finding suggests the need to encourage Bhutanese teachers to use teaching and learning approaches that are responsive to individual students' learning needs during school time.

Discussion

Homework, which is commonly understood as tasks assigned to students by school teachers that are carried out outside of school hours, is reported as a correlate of student performance (Brock, et al., 2007; Cooper, 1989; Cooper, et al., 2006). In line with this, Grade 10 Bhutanese students tended to perform better on the Mathematics test when homework of 15 to 60 minutes was assigned three to four times per week. Similar findings were reported in the OECD (2004a). The existence of certain frequencies and times at which homework contributed to student learning indicates the need to develop homework policies by schools with the aim of fitting homework within the optimal frequency and time.

6.13 Summary and Implications for Policy

This chapter found that students' characteristics such as gender, age, SES, motivation, self-beliefs, anxiety, learning preferences, self-regulation, ICT knowledge and skills, and their views about classroom and school climates, and their experience of homework were correlates of their performance on the Mathematics test. The relationship offered deeper insights into key areas where policy-makers can develop evidence-based policy directives to improve school effectiveness. These insights are elaborated in the following paragraphs.

Disaggregating students by gender found that boys and girls differed in their scores on the Mathematics test, the self-efficacy scale, and the anxiety scale. On average, boys outperformed girls on the Mathematics test, boys had higher self-efficacy about mathematics than girls, and boys had less anxiety about mathematics than girls. These findings highlight the need for interventions for girls to raise their self-efficacy and to reduce their anxiety about mathematics.

Students' age related strongly to their performance on the Mathematics test. Students in the age range of 15 to 17 years performed relatively better on the Mathematics test, though the average age of the Grade 10 students was 18.6 years. Given that the official school enrolment age in Bhutan was six years, the average of 18.6 years suggested Grade retention. As expected, one in every four students reported repeating one or two grades in primary school. These findings show the need for studies on student enrolment-age and Grade retention in the light of the opportunity costs inflicted by such practices upon students and the education system.

Students from high SES families outperformed students from low SES families on the Mathematics test. This signals the need for the Ministry of Education to develop policies on the distribution of educational resources based on students' SES profiles.

Students who had high motivation, high self-efficacy, positive self-concept, and low anxiety about mathematics performed better on the Mathematics test than other students. In addition, students who used self-regulated learning strategies and who preferred a competitive learning environment to a cooperative learning environment performed better than other students. These findings show the need for policy directives that promote motivation, self-efficacy, and a positive self-concept in students. Policy directives may also be aimed at alleviating students' anxiety about mathematics. Similar initiatives may also be focussed on developing flexible learning environments in relation to students' preferences.

Students' knowledge and skills of ICT is as fundamental a determinant as their knowledge and skills of mathematics for successful participation in a globalised economy. About four out of 10 students reported that they had never used computers, a tool fundamental to acquiring ICT knowledge and skills. This group of students may not be able to use ICT as part of their day-to-day activities should they fail to pursue further studies beyond Grade 10, and this may significantly disadvantage them in job opportunities. More than three quarters of students who used computers had access to computers at school, while only about one quarter of them had access to the same at home. This indicates the opportunity for the Ministry of Education to reduce the effects of the digital divide by equipping schools with ICT facilities. Few of the students had been using computers for duration beyond three years. This suggests that perhaps many student users were in higher Grades and used computers, where there was access to ICT facilities, at the cost of the lower Grades. This finding, if proven, would indicate the need for the Ministry of Education to frame policy directives on equitable access to ICT facilities for all students irrespective of their Grades. Fewer than half the students used the Internet to look up information, play online games, keep social contacts, download music and software, and communicate electronically. This indicates a limited popularity of the Internet as a powerful, up-to-date learning and entertainment resource among Grade 10 students. Despite the recent incorporation of ICT into the Bhutanese education system, the majority of students had positive attitudes towards it. This indicates students' willingness to engage with ICT, a learner trait that needs to be rewarded by providing more ICT facilities to schools. Overall, students who used computers performed better than students who did not use them and students who used computers at home and school performed better than students who used them in other places (possibly, a friend's house, Internet Cafe, or a video game parlour). A key area where policy directives may be developed, as conveyed by the findings, is increasing students' access to computers and training at their school.

Students' views about classroom and school climate related to their performance on the Mathematics test. The majority of students experienced great teacher support and few disciplinary problems in the class. Further, teacher support and the classroom disciplinary climate correlated with each other, with high teacher support relating to favourable classroom disciplinary climate. In addition, the majority of students had positive attitudes towards, and a high sense of belongingness to, their school. These findings suggest that the majority of Bhutanese Grade 10 students were engaged in school. In addition, students who had positive attitudes towards school, who felt connected to their school, and who perceived constructive relations with their teachers performed better on the Mathematics test than students who held contrasting attitudes, experiences, and perceptions. These findings point to the need for policy directives aimed at promoting favourable attitudes, attachments, and friendly student and teacher relationships.

The majority of students were assigned homework in mathematics for three or four times per week. The Mathematics test score was optimal for the three or four times per week frequency of homework of 15 to 60 minutes task time. The majority of students reported spending more than three-fifths of their time on homework as compared to the time spent on other learning activities (e.g., remedial classes, enrichment classes, and tutorial classes). Compared to the time spent on doing extended learning activities, more time was spent on the out-of-school-time lessons, and tuition. These findings indicate the need for policy directives that require schools to have homework policies. They also suggest the need for research into why such a heavy out-of-school learning load added to homework should be necessary at all.

6.14 Chapter Conclusion

The chapter showed that the student-related non-cognitive variables as characterized by gender, age, SES, motivation, self-beliefs, anxiety, learning preferences self-regulation, ICT knowledge and skills, and homework were correlates of student achievement in the Mathematics test. A prominent pattern that had emerged from this chapter was that students who identified with high SES, high motivations, favourable self-beliefs, positive attitudes, and used self-regulated learning strategies performed better on the Mathematics test than other students. The findings have a strong potential to assist policy makers at the Ministry of Education in developing a wide range of strategic educational interventions.

Chapter 7

ANALYSES OF THE TEACHER QUESTIONNAIRE

7.1 Introduction

The purposes of this chapter are to report the analyses of data from the teacher questionnaire, and to identify key teacher characteristics that relate to student achievement on the Mathematics test. The results from the analyses are relevant to the research outcome 6 of generating knowledge about teachers and teaching and their effects on student outcomes. Specifically, this chapter will provide responses to the following research questions from Section 4.3.2.1:

ClL 1 How well do gender, age, educational qualification, and teaching experience of teachers relate to student achievement? (Section 7.2)

ClL 2 How well do Bhutanese teachers' demographic profiles compare with those of the teachers across the OECD countries? (Section 7.2)

ClL 3 How well does teacher professional collaboration and development relate to student achievement? (Sections 7.3 and 7.4)

ClL 4 How well do teacher appraisal and feedback relate to student achievement? (Section 7.5)

ClL 5 How well does classroom management relate to student achievement? (Section 7.6)

ClL 6 How well does the teachers' view of their school climate relate to student achievement? (Section 7.7)

ClL 7 How well do teachers' beliefs and teaching practices relate to student achievement? (Sections 7.8 and 7.9)

ClL 8 How well does teachers' self-efficacy relate to student achievement? (Section 7.10)

ClL 9 How well do teachers' usage of homework, tests, textbooks, calculators, and ICT relate to student achievement? (Sections 7.11 through 7.15)

ClL 10 How well do teachers perceive their student engagement? (Section 7.6)

The analyses were done in line with the analytical framework presented in Chapter 5.

7.2 Teachers' Demographic Profiles

Teachers' demographic details were profiled by collecting information about their gender, age, education, and teaching experience. Table 7.1 shows the percentages of students taught by teachers with different demographic profiles, along with their students' mean Mathematics test scores.

Table 7. 1. Percentages of Students Taught by Teachers with Varying Demographic Profiles by Students' Mean Mathematics Test Scores

Teachers' Demographic Profiles			Mathematics Test	
	%	SE	Mean Score	SE
Gender				
Female	16.0	4.93	371.51	9.26
Male	84.0	4.93	358.57	4.55
Age				
Under 25	3.1	2.52	368.44	17.75*
25-29	41.8	7.07	351.15	5.92
30-39	38.0	6.78	365.33	6.80
50-59	17.1	5.50	371.97	9.46
60+				
Academic Degree				
Class 12 Certificate	41.23	7.12	353.48	6.23
Bachelor's Degree	40.45	6.94	369.18	6.76
Master's Degree	18.32	5.10	357.88	7.07
Professional Degree				
B.Ed (Primary)	6.9	3.18	338.74	8.90
B.Ed (Secondary)	62.6	6.91	357.76	5.38
PGCE	24.1	6.09	368.76	7.40
M.Ed	6.4	4.02	381.94	12.45
Mathematics as Academic Major				
Yes	87.2	5.27	363.63	4.31
No	12.8	5.27	340.22	7.52
Grade 10 Teaching Experience				
Two Years	21.69	5.74	351.71	8.39
Three Years	17.59	5.63	355.30	9.90

Four Years	22.33	6.10	360.07	7.48
Five Years plus	38.39	6.84	368.46	7.09

Note: * fewer than 50 students were taught by teachers with this characteristic.

Table 7.1 offers a range of insights into the relationship between the teachers' demographic profiles and the students' performance on the Mathematics test. First, the difference between the mean Mathematics test scores of the students taught by female teachers and of the students taught by male teachers was statistically non-significant, *diff*=12.94, 95% CI [-7.29, 33.16]. Second, the students taught by relatively older teachers performed better on the Mathematics test than the students taught by relatively younger teachers. Especially, the difference between the mean Mathematics test scores of the students taught by the teachers in the age range of 25-29 and in the age range of 50-59 was statistically significant, *diff*=20.82, 95% CI [3.15, 38.50]. Third, the difference in the performances of the students, on the Mathematics test, taught by the teachers with different levels of academic qualifications was statistically non-significant. Fourth, the difference between the mean Mathematics test scores of the students taught by teachers with B.Ed (Primary) and with PGCE qualifications was statistically significant, *diff*=30.02, 95% CI [7.33, 52.71]. Fifth, the difference between the mean Mathematics test scores of the students taught by teachers with an academic major in mathematics and by teachers without an academic major in mathematics was statistically significant, *diff*=23.41, 95% CI [6.43, 40.39]; with the difference being in favour of the students taught by the teachers with a major in mathematics. Sixth, the difference between the mean Mathematics test scores of the students taught by teachers who had been teaching Grade 10 mathematics for two years and by the teachers who had been teaching the same grade for more than five years was statistically non-significant, *diff*=16.75, 95% CI[-4.79, 38.29]. However, a pattern in the relationship between the teachers' years of teaching experience and the students' mathematics test scores seemed to reinforce the common notion that teachers' experience facilitates better student learning.

Discussion

Teachers' demographic profiles were developed by using gender, age, educational attainment, and teaching experience. The analyses of data on the first three of demographic profiles revealed interesting patterns as discussed below.

Gender. Only 16% of Grade 10 students who sat the Mathematics test were taught by female mathematics teachers, indicating a severe gender bias. In other countries, the gender difference in mathematics teachers was mostly insignificant (Mullis, et al., 2008). The analyses of the relationship between the teachers' gender and student performance revealed that the students taught by female teachers tended to outperform the students taught by male teachers. These findings support the reports of other studies that focussed on teachers' gender and student performance (Dee, 2007; Ehrenberg, Goldhaber, & Brewer, 1995; Steele, 1997). These studies noted that teachers' gender influenced student achievement through role modelling, subjective evaluation, negative stereotyping, and classroom interactions. Therefore, appropriate interventions to increase the number of female mathematics teachers in Bhutan are desirable.

Age. Approximately 83% of Grade 10 students who sat the Mathematics test were taught by mathematics teachers who were below 39 years of age, indicating a comparatively young teaching workforce as compared to the percentage of mathematics teachers (51%) in the same age range reported in Mullis, et al. (2008). However, the students taught by relatively older teachers outperformed the students taught by relatively younger teachers, which is consistent with other research findings (Mullis, et al., 2008). The benefits of teachers' experience and budgetary implications from cumulative teaching experience needs a fine balance. As school teachers in Bhutan attain seniority by number of years of service rather than by meritorious performance, having a large number of teachers aged 50 years and above would put pressure on teacher salary and other staff remunerations, a trend emphasized in the OECD (2009a) report from TALIS.

Educational Attainment. Teachers' educational attainments, measured in terms of academic and professional qualifications, showed that both qualifications related to student performance, with higher qualifications resulting in better student performance. Furthermore, the students taught by teachers with a mathematics major outperformed the students taught by teachers without the same. While these findings are in line with reports from other studies (Barber & Mourshed, 2007), their values in planning professional development programmes for teachers need special emphasis in the Bhutanese education system. For instance, three-fifths of Grade 10 students were taught by teachers with a B.Ed degree, while only one quarter of students in the same grade were taught by teachers with a PGCE qualification. The OECD (2001, p. 11) notes that "teachers' subject-matter expertise must be complemented by pedagogical competence", which in the Bhutanese education system might mean providing academic professional development programmes to teachers with B.Ed qualifications and, additionally, pedagogical support to teachers with PGCE qualifications. Therefore, the findings on teacher educational attainment indicate the need to provide strategic professional development programmes to all teachers.

7.3 Teachers' Professional Collaboration

Teachers were asked to report on the activities related to professional collaboration. As shown in Table 7.2, the majority of students (69.6%) were taught by teachers who collaborated with other teachers and discussed the teaching of mathematics concepts. Many students, however, were taught by teachers who never had their class observed by other teachers (52.2%) or who never observed other teachers' classes (43.5%).

Table 7. 2. Percentages of Students by Their Teachers' Frequency of Professional Collaboration with Other Teachers

Activities	Never or almost never		2 or 3 times per month		At least Weekly	
	%	SE	%	SE	%	SE
Discussion about how to teach a particular concept			30.4	6.66	69.6	7.43
Working on preparing instructional materials	5.1	3.81	42.2	6.94	52.6	8.89
Visits to another teacher's classroom to observe his/her teaching	43.5	7.04	45.8	7.07	10.6	3.53
Informal observation of my classroom by another teacher	52.2	7.10	47.8	7.10		

The relationship between the index of the frequency of teachers' professional collaboration (TPC) and their students' Mathematics test scores are shown in Table 7.3.

Table 7. 3. Students' Mean Mathematics Test Scores by Quartiles of TPC

Quatile	Teachers' Professional Collaboration		Mathematics Test	
	Mean Score	SE	Mean Score	SE
1	-1.85	0.17	361.08	6.18
2	-.69	0.07	360.43	7.57
3	.30	0.07	355.64	6.69
4	1.03	0.02	365.40	11.29

The difference in the mean Mathematics test scores of the students in the bottom and the top quartiles TPC was statistically non-significant, *diff*=4.32, 95% CI [-20.92, 29.55], showing that the students' learning was not impacted upon by their teachers' frequency of professional collaboration.

Discussion

The majority of students were taught by teachers who reported discussing how to teach certain mathematical concepts and working together on instructional materials on a weekly basis, while approximately one-half of the students were taught by teachers who reported never observing another teacher's class or having their class observed by other teachers. These findings are consistent with the findings reported in the OECD (2009a), indicating teachers' lack of involvement in team teaching and observing colleagues' classes. In addition, it was likely that the teachers worked and discussed teaching approaches superficially without referring to research-based information (OECD, 2009a). The plausibility of this view is evident in the absence of a difference in the performances of the students on the Mathematics test taught by teachers who reported high engagement in professional collaborations and the students taught by teachers who reported low engagement.

7.4 Professional Development

In this section teachers' professional development is analysed to understand its position in the Bhutanese educational system by looking at data related to teachers' participation in it, areas of teachers' needs for it, and reasons given by teachers' for not participating in it.

7.4.1 Teachers' Participation in Professional Development

Teachers were asked to indicate various professional development programmes in which they had participated in the past two years of their teaching career. As shown in Table 7.4, more than 50% of the students had mathematics teachers who did not participate in any professional development programmes. Further, none of the students had mathematics teachers who observed the classes of mathematics teachers in different schools.

Table 7. 4. Percentages of Students by Their Teachers' Participation in Professional Development

Professional Development Programs	Participated		Not Participated	
	%	SE	%	SE
Mathematics content	26.8	6.21	73.1	6.20
Mathematics pedagogy/instruction	19.3	5.88	80.7	5.88
Mathematics curriculum	45.2	7.03	54.8	7.03
Integrating technology into mathematics	13.4	4.92	86.6	4.92
Improving students' critical thinking or problem solving skills	18.9	5.79	81.1	5.79
Mathematics assessment	24.4	5.90	75.6	5.90

Effective use of manipulatives in mathematics instructions	13.4	4.75	86.6	4.75
Effective use of calculators in mathematics instructions	13.5	5.14	86.5	5.14
College course taken after your first certification	6.0	2.98	94.0	2.98
Conference or professional association meeting	10.6	4.27	89.4	4.27
Observational visit to another school			100.0	0.00
Mentoring and/ or peer observation and coaching as part of a formal school arrangement	44.4	7.09	55.6	7.09
Committee or task force focussing on curriculum, instruction, or student assessment	30.6	6.48	69.4	6.48
Regularly scheduled discussion or study group	27.4	6.06	72.6	6.06
Teacher collaborative or network	21.7	6.20	78.3	6.20
Individual or collaborative research	19.9	5.39	80.1	5.39
Independent reading on a regular basis	68.0	6.63	32.0	6.63
Co-teaching or team teaching	17.2	5.08	82.8	5.08
Consultation with a mathematics specialist	23.6	5.91	76.4	5.91

7.4.2 Teachers' Professional Development Needs

Teachers were asked to indicate the areas where they felt they needed professional development programmes. As shown in Table 7.5, more than 50% of students were taught by teachers who reported a moderate- to high-level need for most of the professional development programmes in all areas of need except those involving classroom management.

Table 7. 5. Percentages of Students Taught by Teachers Needing Professional Development

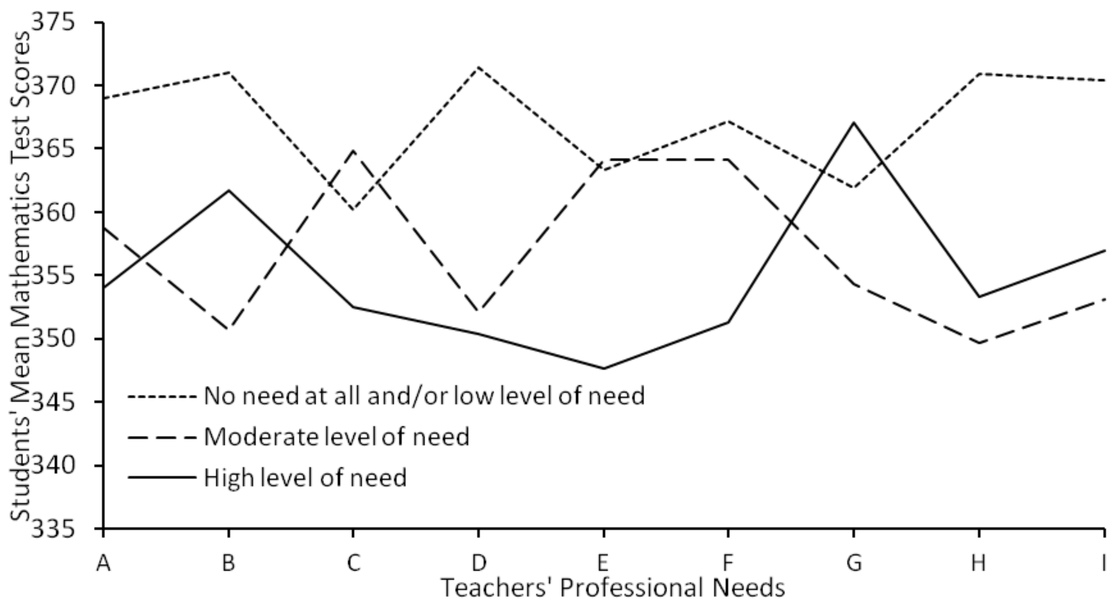
Need Areas	No need at all		Low level of need		Moderate level of need		High level of need	
	%	SE	%	SE	%	SE	%	SE
Content and performance standards of mathematics	2.8	2.06	31.6	6.81	31.4	6.55	34.2	6.66
Student assessment practices	9.7	4.04	30.4	6.46	43.5	7.07	16.3	5.48
Classroom management	30.0	6.31	33.7	6.86	26.0	6.31	10.2	4.27

Knowledge and understanding of mathematics	22.8	5.81	23.4	6.10	31.6	6.69	22.1	5.88
Knowledge and understanding of instructional practices in mathematics	12.0	4.67	31.6	6.87	37.3	6.73	19.0	5.49
ICT skills for teaching	8.5	3.84	24.8	6.10	31.5	6.73	35.1	6.75
Student disciplines and behavioural problems	29.1	6.41	37.3	6.87	23.8	6.14	9.7	4.27
School management and instructional problems	22.2	6.21	26.5	6.15	34.4	6.75	16.9	5.12
Student counselling	8.6	3.62	30.0	6.50	39.7	7.07	21.7	5.70

As high as 97.2% of the students were taught by teachers who had at least a low-level need for a professional development programme on the content and performance standards of the Grade 10 Mathematics Curriculum. Similarly, 91.5% of the students were taught by teachers who had at least a low-level need for a professional development programme on using ICT in teaching.

The relationship between the teachers' professional development needs and students' performance on the Mathematics test was presented in Figure 7.1. As shown in Figure 7.1, the students taught by teachers who reported no need or a low level of need for professional development programmes scored relatively higher than the students taught by teachers who reported a moderate or a high level of need for professional development programmes.

A pattern visible in Figure 7.1 (where a line graph is used for categorical data to emphasise the patterns which are not obvious with a column graph) is the section of the "High level of need" line from C to F. This section of the line indicates potential starting points for designing strategic educational interventions.



Note:
A=Content and performance standards of mathematics; B=Student asessment practices;
C= Classroom management; D= Knowledge and understanding of mathematics;
E=Knowledge and understanding of instructional proctices in mathematics;
F=ICT skills for teaching; G=Student disciplines and behavioural problems;
H=School management and instructional problems; and I=Student counselling.

Figure 7. 1. Teachers' Professional Needs by Students' Mean Mathematics Test Scores

7.4.3 Teachers’ Reasons for Not Participating in Professional Development

As shown in Table 7.6, among different reasons that prevented teachers from participating in professional development programmes, the lack of a suitable professional development programme was the strongest reason.

Table 7. 6. Teachers' Reasons for Not Participating in Professional Development Programmes

Reasons	%	SE
I did not have the pre-requisites	12.8	4.79
Professional development was too expensive	1.7	0.24
There was a lack of school support	1.3	0.18
Professional development conflicted with my work schedule	27.3	6.48
I did not have time because of family responsibilities	5.7	3.38
There was no suitable professional development programme offered	51.3	7.12

Discussion

Analyses of data on different aspects of teacher professional development, as related to Grade 10 mathematics teachers in Bhutan, showed that over one-half of the students were taught by teachers who did not participate in any professional development in the past two years, while over one-half the students were taught by teachers who reported moderate to high-level needs for professional development. Further, the students taught by teachers who reported a moderate- to high-level need for professional development programmes, on average, performed at a lower level on the Mathematics test than the students taught by teachers who reported either no need or a low level of need for professional development. This shows a negative relation between teachers' professional development needs and student performance. Yoon, Duncan, Lee, Scarloss, and Shapley (2007), in their review of 1,300 studies on the effect of teacher professional development on student achievement, reported an average effect size of 0.54. A sustained opportunity for Bhutanese teachers to participate in professional development is desirable because the majority of the Grade 10 students in Bhutan were taught by teachers with a B.Ed qualification. Also, the majority of teachers reported a moderate- to a high-level need for professional development. The need for strategic professional development programmes for Grade 10 mathematics teachers is evident in their reasons for not having participated in professional development in the past two years of their teaching career, with over one-half of the teachers citing the lack of suitable professional development as the main reason.

7.5 Teachers' Appraisal

This section examines the level of emphasis made by school principals on different areas of teacher appraisal. As Table 7.7 indicates, students were taught by teachers who reported varying degrees of emphases made by their principals during their appraisals.

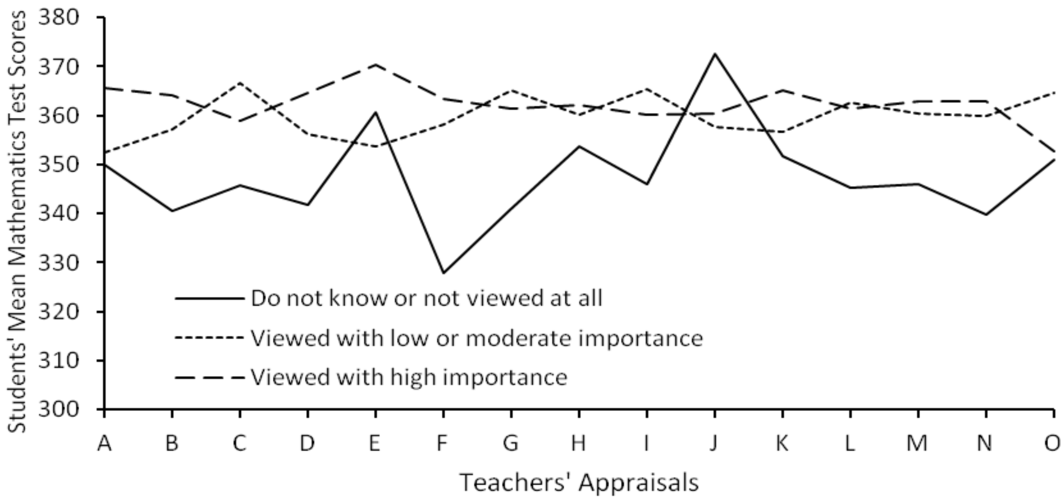
Table 7. 7. Percentages of Students Whose Teachers Reported Being Appraised in Various Areas by School Principals

Appraisal Focus	Do not know		Not viewed at all		Viewed with low importance		Viewed with moderate importance		Viewed with high importance	
	%	SE	%	SE	%	SE	%	SE	%	SE
Student test scores	2.0	0.28	4.2	3.03	4.8	2.73	25.6	6.19	63.4	6.82
Retention and pass rates of students			3.6	2.55	1.2	0.17	36.9	6.84	58.3	6.98
Other student learning outcomes	3.9	0.53	7.9	3.89	3.4	2.57	39.0	7.11	45.8	6.97

Student feedback on my teaching	1.2	0.17	6.3	3.10	5.3	3.01	22.0	6.20	65.3	6.82
Feedback from parents	14.4	4.78	10.8	4.05	6.2	3.11	37.2	7.00	31.4	6.68
How well I work with the Principal and my colleagues	1.6	0.23	1.7	0.24	1.0	0.14	30.0	6.52	65.8	6.70
Direct appraisal of my classroom teaching	1.6	0.23	7.4	3.30	1.9	0.27	26.1	6.03	63.0	6.68
Innovative teaching practices	2.6	0.36	6.4	3.17	2.9	2.15	37.3	6.85	50.8	7.11
Relations with students	5.8	2.89	1.7	0.24	1.9	0.27	28.1	6.56	62.5	6.91
Professional development I have undertaken	3.5	2.48	6.2	3.77	6.7	3.32	27.9	6.42	55.7	7.08
Classroom management			3.4	2.37	4.6	2.64	43.2	7.19	48.8	7.09
Knowledge and understanding of my main subject	1.3	0.18	5.0	2.85	3.3	2.33	13.9	4.59	76.6	5.66
Knowledge and understanding of instructional practices	1.2	0.17	6.3	3.11	3.6	2.56	37.2	7.10	51.7	7.12
Student discipline and behaviour	1.3	0.18	1.7	0.24			52.7	7.10	44.3	7.07
Extra-curricular activities with students	3.8	2.84	1.7	0.24	11.4	4.36	56.1	6.93	27.0	5.82

It is interesting to note that about 14% of the students were taught by teachers who reported not knowing whether their principals considered parents' feedback as part of their appraisal. About 11% of the students were taught by teachers who reported that the parents' feedback was not an element of their appraisal.

Figure 7.2 shows the relation between students' performance on the Mathematics test and their teachers' appraisal focus (A line graph is used for categorical data to emphasise the patterns which are not obvious with a column graph). Figure 7.2 shows that the students taught by teachers who reported that their principals viewed the various areas of appraisals with low, moderate, or high importance, on average, performed at a higher level on the Mathematics test than the students taught by teachers who reported that they were not aware of the appraisal or were not viewed at all.



Note:
A=Student test scores; B=Retention and pass rates of students;
C=Other student learning outcomes; D=Student feedback on my teaching;
E=Feedback from parents; F=How well I work with the Principal and my colleagues;
G=Direct appraisal of my classroom teaching; H=Innovative teaching practices;
I=Relations with students; J=Professional development I have undertaken;
K=Classroom management; L=Knowledge and understanding of my main subject;
M=Knowledge and understanding of instructional practices;
N=Student discipline and behaviour; and O=Extra-curricular activities with students.

Figure 7. 2. Teachers' Appraisal Areas by Students' Mean Mathematics Test Scores

Discussion

The OECD (2009a) notes that teacher appraisal increases teachers' job satisfaction, job security, and development as teachers. The analyses of these data on teacher appraisal revealed that the majority of students were taught by teachers who were aware of the areas in which they were appraised by their principals, indicating proper communication between teachers and principals. The majority of students were taught by teachers who reported that the following were viewed with high importance by their principals: student test scores; retention and pass rates of students; student learning outcomes; collegial relationship; teaching practices; participation in professional development; classroom management; subject knowledge; pedagogical knowledge; student discipline; and extra-curricular activities. On the other hand, approximately one-quarter of the students were taught by teachers who reported that they were not aware if their principals used parents' feedback or their principals did not use parents' feedback on their teaching to appraise them, indicating an absence of teachers' accountability to parents. The analysis also revealed that the students of teachers who reported that their principals viewed the appraisal areas with low, moderate, or high importance performed relatively higher on the Mathematics test than the students of teachers who reported that they were not aware of the appraisal or were not viewed at all, which supports similar claims made in the OECD (2009a).

7.6 Teachers' View of Classroom Climate

Teachers were asked to report about the classroom climate. Table 7.8 indicates that fewer than 18% of the students were taught by teachers who experienced some classroom problems. On the contrary, over 89% of the students were taught by the teachers who experienced students' efforts to create a pleasant learning atmosphere in the class.

Table 7. 8. Percentages of Students Whose Teachers' Reported Agreeing or Strongly Agreeing with Classroom Climate Variables

Measures	%	SE
When the lesson begins, I have to wait quite a long time for students to pay attention	13.2	4.11
Students in this class take care to create a pleasant learning atmosphere	89.1	8.69
I lose quite a lot of time because of students interrupting the lesson	16.7	4.56
There is much noise in the classroom	17.2	5.35
Students do not know about my expectations for their classroom behaviour	17.1	5.05

Table 7.9 indicates the relation between the quartiles of the classroom climate index (TCDC, this index has been explained in Section 4.6.2.2.3 of Chapter 4) and the students' mean Mathematics test scores, with the top two quartiles of the index of classroom climate corresponding to higher Mean Mathematics test scores.

Table 7. 9. Quartiles of TCDC by Students' Mean Mathematics Test Scores

Quartiles	TCDC		Mathematics Test	
	Mean	SE	Mean	SE
1	-0.49	0.10	353.11	7.72
2	0.50	0.06	352.57	6.88
3	0.95	0.09	373.37	10.27
4	2.92	0.19	365.76	6.70

Discussion

The majority of students were taught by teachers who reported a positive classroom climate, indicating that teachers viewed their students as engaged in school. However, the fact that almost one-fourth of the students were taught by teachers who reported disciplinary problems in the classroom indicates the need for the teachers to implement appropriate measures to improve classroom climate. Such measures may involve setting classroom rules, developing disciplinary interventions, building supportive teacher-student relationships, and building an appropriate mental set. These four elements are frequently linked to positive classroom climate (Marzano, 2003a, 2003b).

7.7 Teachers' View of School Climate

Overall, the majority of students were taught by teachers who had a positive disposition toward their school. As Table 7.10 shows, except for the parental support for student activities (47%), a maximum of 19% of students were taught by teachers who reported that their students had low or very low regard for school property.

Table 7. 10. Percentages of Students Whose Teachers Reported Their Perspectives about Different Variables of School Climate

Elements of School Climate	Very high		High		Medium		Low		Very Low	
	%	SE	%	SE	%	SE	%	SE	%	SE
Teachers' job satisfaction	17.5	5.18	42.4	7.10	28.0	6.15	12.0	4.92		
Teachers' understanding of the school's curricular goal	24.5	5.97	65.8	6.51	9.6	3.50				

Teachers' degree of success in implementing the school's curriculum	18.7	5.17	68.5	6.60	12.8	5.22				
Teachers' expectation for student achievement	39.4	6.85	34.6	6.85	23.5	6.25	2.4	1.73		
Parental support for student activities	3.8	2.40	13.3	4.37	35.0	6.87	29.0	6.61	18.7	5.53
Students' regard for school property	4.3	2.54	14.9	5.13	62.1	6.81	10.0	3.97	8.7	3.89
Students' desire to do well in school	10.0	4.04	28.4	6.38	53.2	7.07	7.5	3.41	0.9	0.13

Table 7.11 shows the quartiles of the index of teachers' view of school climate (TSCHL, this index has been explained in Section 4.6.2.2.3 of Chapter 4) with the mean mathematics test scores.

Table 7. 11. Quartiles of TSCHL by Students' Mean Mathematics Test Scores

Quartiles	TSCHL		Mathematics Test	
	Mean	SE	Mean	SE
1	-1.31	0.19	359.55	5.50
2	-0.09	0.07	366.57	10.03
3	0.62	0.07	362.07	6.24
4	1.85	0.13	356.11	9.14

As shown in Table 7.11, teachers' positive views of the school climate did not show a clear pattern in the mean Mathematics test scores. However, the mean Mathematics test scores were higher in the second and the third quartiles of TSCHL.

Discussion

The majority of teachers identified themselves as working in a positive school climate. Most of the students were taught by teachers who rated high to very high prevalence of their understanding of school curricular goals, success rate in implementing school curriculum, high expectations for student achievement, parental support for student activities, students' regard for school properties, and students' desire to do well in their schools. Such positive aspects are indicative of an orderly school climate (Marzano, 2001; Scheerens, et al., 2003).

However, there was no discernible pattern in the Mathematics test scores of the students taught by the teachers in the top and the bottom quartiles of the index of school climate. This finding differs from the positive relationship between school climate and student achievement reported in the literature (Brand, et al., 2003; Hoy & Hannum, 1997; Macneil, et al., 2007; Mullis, et al., 2008; OECD, 2004a). The lack of association between school climate and student achievement in the Bhutanese education system may be attributed to teachers' inclination to identify with what they deemed as desirable characteristics of school climate when responding to the teacher questionnaire. The lack of association between school climate and student achievement may also be attributed to a less interactive experience of the teachers with the factors of the school climate.

7.8 Teachers' Beliefs About Teaching Approaches

Teachers were asked to indicate their beliefs in two broad teaching approaches, namely, direct transmission and constructivist approaches. As shown in Table 7.12, overall, the percentages of students taught by teachers who believed in constructivist approaches was higher than the percentages of students taught by teachers who believed in direct transmission approaches.

Table 7. 12. Percentages of Students Taught by Teachers Believing in Direct Transmission or Constructivist Approaches to Teaching

Beliefs	%	SE
Direct Transmission		
Mathematics should be learned as sets of algorithms or rules that cover all possibilities	62.7	8.28
Learning mathematics mainly involves memorizing	10.6	5.04
Few new discoveries in mathematics are being made	82.4	6.51
Effective or good teachers demonstrate the correct way to solve a problem	93.5	9.92
It is better when the teacher, not the student decides what activities are to be done	68.3	7.98
Teachers know a lot more than students; they should not let students develop answers that may be incorrect when they can just explain the answers directly	28.5	7.23
Instruction should be built around problems with clear, correct answers, and around the ideas that most students can grasp quickly	96.9	9.37

A quiet classroom is generally needed for effective learning	50.1	8.17
Constructivist		
Modelling real-world problems is essential to teaching mathematics	100.0	10.06
More than one representation (picture, concrete material, symbols, etc.,) should be used in teaching a mathematics topic	100.0	10.04
My role as a teacher is to facilitate students' own inquiry	86.0	9.12
Solving mathematics problems often involves hypothesizing, estimating, testing, and modifying findings	95.5	9.97
Students learn best by finding solutions to problems on their own	89.7	9.92
There are different ways to solve most mathematical problems	100.0	9.99
How much students learn depends on how much background knowledge they have-that is why teaching facts is not necessary	48.6	8.58
Students should be allowed to think of solutions to practical problems themselves before the teacher shows them how they are solved	98.9	9.02
Thinking and reasoning processes are more important than specific curriculum learning	91.1	9.53

As shown in Table 7.13 below, teachers' positive belief in direct transmission teaching approaches (TDTM, this index has been explained in Section 4.6.2.2.3 of Chapter 4) seemed to associate with increasing student test scores. On the contrary, teachers' positive belief in the constructivist teaching approach (TDCT, this index has been explained in Section 4.6.2.2.3 of Chapter 4) seemed to associate with decreasing student test scores. However, neither TDTM nor TDCT showed statistically significant associations with student test scores.

Table 7. 13. Quartiles of TDTM and TDCT by Students' Mean Mathematics Test Scores

Quartile	TDTM		Mathematics Test Scores		TDCT		Mathematics Test Scores	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1	-0.02	0.02	359.40	6.16	-0.87	0.10	360.35	6.20
2	0.48	0.04	351.82	7.09	0.00	0.04	363.43	7.52
3	1.09	0.04	362.28	7.94	0.47	0.06	363.95	8.29
4	1.79	0.09	371.77	8.68	1.52	0.16	354.84	7.73

Discussion

In line with the OECD (2009a) report, this study also categorized teachers' beliefs as transmissive or constructivist. The analyses of data on these two belief categories of teachers revealed that more of the teachers identified favourably with constructivist views of teaching and learning than with direct transmission views of teaching and learning. Because teachers' beliefs about the nature of teaching significantly influence their choice of classroom practices (OECD, 2009a), teachers in Bhutan appear to be using constructivist teaching approaches more frequently than the transmissive teaching approaches. However, the ways in which the constructivist teaching approaches are used in the classroom needs further study because students appeared to be achieving higher average test scores for teachers who reported a preference for transmissive teaching approaches. An alternative way of interpreting this finding is that the teachers might have responded to the questionnaire as asking what they perceived as desirable practices; therefore, this finding needs to be interpreted and acted on with caution.

7.9 Teachers' Use of Teaching Strategies

Teachers' use of structured, constructivist and extended teaching strategies were analysed in this section. In addition, teachers' use of reinforcement strategies as a component of teaching practices was also analysed in this section.

7.9.1 Structured, Constructivist, and Extended Teaching Strategies

Table 7.14 shows the percentages distributions of the students taught by teachers who practised structured, constructivist, and extended teaching strategies in the classroom.

Table 7. 14. Percentages of Students Whose Teachers Reported Practising Structured, Constructivist, and Extended Teaching Strategies

Teaching Strategy	Never or hardly ever		In one-quarter of lessons		In about half of lessons		In three quarters of lessons		In almost every lessons	
	%	SE	%	SE	%	SE	%	SE	%	SE
Structured Teaching										
Presenting a short summary of the previous lesson at the beginning of a new lesson			13.9	5.40	6.1	3.47	13.0	4.58	67.0	6.81
Stating learning goals	2.1	1.53	27.5	6.70	17.8	5.72	7.3	3.35	45.2	7.01
Students listening to lecture-style presentations	13.2	4.73	34.9	6.80	23.9	5.88	16.3	5.53	11.7	4.60

Students working problems with your guidance			9.8	4.76	29.7	6.79	18.3	4.87	42.2	6.97
Asking students to remember every step in a procedure	10.7	4.55	25.6	6.36	18.5	5.29	13.4	5.18	31.7	6.45
Students listening to you re-teach and clarify content/procedure	0.8	.12	17.7	5.01	27.8	6.63	18.6	5.68	35.0	6.74
Asking questions to students on the lesson contents			15.4	5.32	9.7	4.79	9.6	3.73	65.3	6.88
Giving notes to students	8.3	3.68	15.5	4.97	16.3	5.22	16.0	5.39	43.9	7.08
Checking students' exercise books			14.0	4.94	22.5	5.81	22.3	6.14	41.2	6.98
Summarising the lesson contents			14.5	5.20	9.7	4.50	25.1	6.35	50.7	7.13
Administering tests or quizzes	6.7	4.11	32.8	6.48	20.5	5.74	16.1	5.55	23.9	5.97
Reviewing students' homework			10.3	4.35	16.8	5.16	14.6	5.44	58.4	7.06
Constructivist Teaching										
Planning classroom activities with students	15.3	5.63	23.2	5.92	20.1	5.87	16.3	4.98	25.2	5.97
Students working in small groups on a common problem	3.2	.44	25.0	6.37	30.5	6.65	21.3	5.65	19.9	5.17
Working with individual students	2.6	1.87	32.4	6.98	21.4	6.03	15.7	4.55	28.0	6.17
Students interpreting data in tables, charts, or graphs	1.0	.14	39.3	7.02	20.8	5.65	25.9	6.30	13.1	4.59
Students writing equations and functions to represent relationships	1.7	.24	18.6	5.49	37.5	6.94	23.9	6.17	18.3	5.26
Students explaining their answers	1.7	.24	30.4	6.67	28.4	6.39	11.1	4.23	28.4	6.40
Students deciding on their own procedures for solving	9.7	4.74	41.3	7.04	21.9	5.50	14.7	5.20	12.4	4.41

complex problems										
Students comparing lesson contents or topics or methods	4.7	2.70	31.0	6.62	21.4	5.82	24.5	6.38	18.3	5.19
Students guessing and predicting solutions to problems	2.4	1.70	22.6	5.95	21.0	6.00	24.4	6.25	29.6	6.32
Students working problems on their own without your guidance			33.2	6.68	26.6	6.02	12.7	5.15	27.4	6.44
Extended Teaching										
Students working on problems for which there is no immediately obvious method of solution	24.9	5.95	43.6	7.13	11.0	4.29	14.1	4.73	6.4	3.66
Students relating what they are learning in mathematics to their daily lives	1.1	.15	26.7	6.07	20.2	6.03	23.8	6.11	28.2	6.40
Students working on projects that require at one week to complete	26.6	6.64	37.5	6.85	19.9	5.56	4.5	2.49	11.5	4.34
Students participating in classroom management tasks not related to the lesson's content	33.7	6.86	38.8	6.97	15.8	5.14	5.5	2.86	6.3	2.82

In Table 7.14, the percentages of students taught by teachers who practised different aspects of a structured-teaching strategy in almost every lesson were generally higher than the percentages of students taught by teachers who practised constructivist-teaching strategies in almost every lesson. The percentages of the students taught by teachers who practised different aspects of extended-teaching strategy were smaller than the percentages of students taught by teachers who practised either structured or constructivist teaching strategies.

Table 7.15 shows the quartiles of the indices of structured teaching (CTPST), constructivist teaching (CTPSO), and extended teaching (CTPEL), together with the students' mean Mathematics test scores. These indices have been explained in Section 4.6.2.2.3 of Chapter 4

Table 7. 15. Quartiles of CTPST, CTPSO, and CTPEL by Students' Mean Mathematics Test Scores

Q	CTPST		Math Test Scores		CTPSO		Math Test Scores		CTPEL		Math Test Scores	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1	-0.06	0.01	351.40	2.19	-0.49	0.10	352.50	0.49	-1.81	0.23	356.93	7.74
2	0.35	0.01	375.02	4.18	0.18	0.03	370.99	0.18	-0.51	0.06	367.48	9.86
3	0.70	0.01	354.81	3.07	0.71	0.04	359.06	0.71	0.30	0.06	358.40	5.94
4	1.43	0.02	361.32	2.73	1.31	0.12	359.84	1.31	1.55	0.27	359.73	5.15

Q=Quartile

As shown in Table 7.15, the relationship of three indices of teaching strategies, namely, structured teaching (CTPST), constructivist teaching (CTPSO), and extended teaching (CTPEL) to students' test scores did not reveal any predictable association when the indices were considered separately. However, students' mean mathematics test scores were comparatively consistent across the fourth quartile of CTPST, CTPSO, and CTPEL, as shown in Table 7.15. In addition, the mean mathematics test scores show a peak in the second quartile across all three indices of teaching strategies.

7.9.2 Teachers' Use of Reinforcement Strategies

Teachers were asked how frequently they used different types of reinforcement strategies in the classroom. Table 7.16 shows the percentages of students taught by the teachers who reported using different reinforcement strategies.

Table 7. 16. Percentages of Students Whose Teachers Reported their Use of Reinforcement Strategies

Reinforcements	Never or hardly ever		In one-quarter of lessons		In about half of lessons		In three quarters of lessons		In almost every lesson	
	%	SE	%	SE	%	SE	%	SE	%	SE
Narrating episodes related to role of efforts in improving achievement	6.6	2.99	41.6	7.15	24.8	5.96	5.8	3.30	21.2	5.68
Asking students to describe relationship between their effort and achievement	9.6	4.15	40.3	7.13	13.9	4.50	12.5	4.49	23.9	5.94

Using concrete symbols to reinforce and recognise students' achievements	8.8	3.62	15.7	5.28	22.8	6.31	29.1	6.30	23.6	6.01
Displaying students work in class	6.7	3.13	38.7	7.09	18.5	5.19	13.2	4.52	22.8	6.12

As shown in Table 7.16, the majority of the students were taught by teachers who reported using different aspects of the reinforcement strategy at least in one-quarter of their lessons. However, less than one-tenth of the students were taught by teachers who reported never using particular reinforcement strategies in the classroom.

As shown in Table 7.17 below, students' mean Mathematics test scores corresponding to the quartiles of the index of teachers' use of reinforcement strategies (TREM, this index has been explained in Section 4.6.2.2.3 of Chapter 4) were not very different from each other, indicating the absence of the effect of reinforcements on student achievement on the Mathematics test.

Table 7. 17. Quartiles of TREM by Students' Mean Mathematics Test Scores

Quartiles	TREM		Mathematics Test	
	Mean	SE	Mean	SE
1	-1.07	0.12	361.18	3.15
2	-0.22	0.03	363.25	3.61
3	0.63	0.02	356.31	3.99
4	1.97	0.07	361.82	2.04

Discussion

Analyses of data on teaching practices of Grade 10 mathematics teachers showed that the teachers used multiple teaching strategies to teach mathematics, which is commendable because teachers need to use a range of teaching strategies to cater for the different learning needs of students (OECD, 2009a). Specifically, the majority of students were taught by teachers who reported using structured teaching strategies most frequently as compared to their report of using constructivist and extended teaching strategies. Comparing the students' mathematics test scores across the top quarters of three indices of teaching strategies, the students taught by the teachers using a structured teaching strategy performed at a higher level on the Mathematics test than the students taught by teachers who reported using constructivist or extended teaching strategies. This trend is markedly visible in the second quartile across three indices of teaching strategies. The second quartile also shows a peak in students' mean mathematics test scores across all three indices of teaching strategies which is worthy of further research.

Reinforcement has been known to improve students' classroom behaviour (Stage & Quiroz, 1997) and student achievement (Fraser, et al., 1987). In the light of such an important role of reinforcement in the teaching and learning process, it is encouraging that more than one-half of the teachers reported using various reinforcement strategies in their class. However, the teachers' report of using various reinforcement strategies did not result in significant improvement in student performance. Therefore, observational studies may have to be pursued to validate the teachers' claim that they use reinforcement strategies in their classes.

7.9.3 Teachers' Views of Factors Constraining Effective Teaching

Teachers were asked to report how their classroom teaching was affected by factors related to student characteristics and material resources on a Likert-type scale, with response options of "Not applicable", "Not at all", "A little", "Some", and "A lot". Table 7.18 shows in part the results.

Table 7. 18. Percentages of Students Whose Teachers Reported Their Ability to Teach Effectively was Constrained by Student- and Resource-related Factors

Constraints	Some		A lot	
	%	SE	%	SE
Student-related				
Students with different academic abilities	29.6	6.57	37.7	6.87
Students who come from a wide range of backgrounds	40.9	7.11	16.5	4.75
Students with special needs	11.2	4.65	27.2	5.76
Uninterested students	21.4	5.87	21.9	5.59
Low morale among students	29.6	6.41	16.6	4.98
Disruptive students	30.8	6.48	6.1	2.69
Resource-related				
Shortage of computer hardware	13.2	4.90	11.3	4.33
Shortage of computer software	15.3	5.53	14.3	4.52
Shortage of support for using computers	19.8	5.59	11.8	4.69

Shortage of textbooks for students' use	26.1	6.53	9.6	3.43
Shortage of other instructional equipment for students' use	24.4	6.04	21.1	5.40
Shortage of equipment for use in demonstrations and other exercises	25.0	5.90	25.8	6.10
Inadequate physical facilities	26.1	6.24	8.9	3.68
High student/teacher ratio	25.5	6.06	31.9	6.54

As shown in Table 7.18, more than 50% of the students were taught by teachers who reported that their ability to teach effectively was limited in some or a lot of ways by different academic abilities and backgrounds of students or a shortage of equipment for classroom demonstrations, or a high student to teacher ratio.

To study the association of the factors constraining effective teaching and student achievement, the quartiles of the indices of student-related constraints (TCONST, this index has been explained in Section 4.6.2.2.3 of Chapter 4), resource-related constraints (TCONRT, this index has been explained in Section 4.6.2.2.3 of Chapter 4) and the corresponding mean Mathematics test scores were tabulated. Table 7.19 shows the result.

Table 7. 19. Quartiles of TCONST and TCONRT by Students' Mean Mathematics Test Scores

Quartile	TCONST		Mathematics Test Scores		TCONRT		Mathematics Test Scores	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1	-0.60	0.14	355.66	9.32	-1.36	0.47	353.59	7.36
2	0.15	0.03	362.00	10.53	-0.03	0.02	361.06	10.61
3	0.63	0.05	366.67	5.84	0.31	0.03	367.99	7.48
4	1.72	0.20	358.22	5.80	1.04	0.12	359.89	6.61

As shown in Table 7.19, neither the student-related constraints nor the resource-related constraints showed any discernible pattern in the students' mean performance on the Mathematics test, indicating a lack of association between the constraints and student achievement.

Discussion

The analyses showed that teachers' ability to teach effectively was somewhat constrained by both student- and resource-related constraints. However, the absence of association between the severity of the reporting constraints and student achievement, as revealed by Table 7.19, may indicate that the teachers did not differ in the ways they taught their students with or without the constraints (Greenwald, et al., 1996; Hanushek, 1997). This finding implies either that teachers who faced the constraints used alternative measures or teachers who did not face the constraints did not make any particular advantage of the lack of the constraints. This indicates a need for an independent study into the usage of educational resources by schools.

7.10 Teachers' Self-Efficacy

Data on teachers' self-efficacy were analysed in terms of the teachers' readiness to teach different topics of the Grade 10 Bhutanese mathematics curriculum and their views about their ability to teach mathematics to Grade 10 Bhutanese students.

7.10.1 Teachers' Readiness to Teach

Teachers were asked how prepared they were to teach seven major topics in the Grade 10 Mathematics Curriculum. As Table 7.20 shows, at least two-thirds of the students were taught by teachers who claimed that they were very ready to teach all the seven topics.

Table 7. 20. Percentages of Students Whose Teachers Reported Feeling Ready or Very Ready to Teach Grade 10 Mathematics Topics

Subject Topics	Ready		Very Ready	
	%	SE	%	SE
Matrices and networks	19	6.18	81	6.18
Linear functions and relations	27.9	6.83	72.1	6.83
Measurement	32	6.82	68	6.82
Quadratic and absolute value functions	32.7	6.99	67.3	6.99
Data management, statistics, and probability	27.8	6.66	72.2	6.66
Trigonometry	25.7	6.74	74.3	6.74
Geometry	19	6.18	81	6.18

As shown in Figure 7.3 (where a line graph is used for the categorical data to emphasise the patterns which are not obvious in a column graph.), the students taught by teachers who reported being very ready to teach scored higher average Mathematics test scores than the students who were taught by teachers who reported being ready to teach.

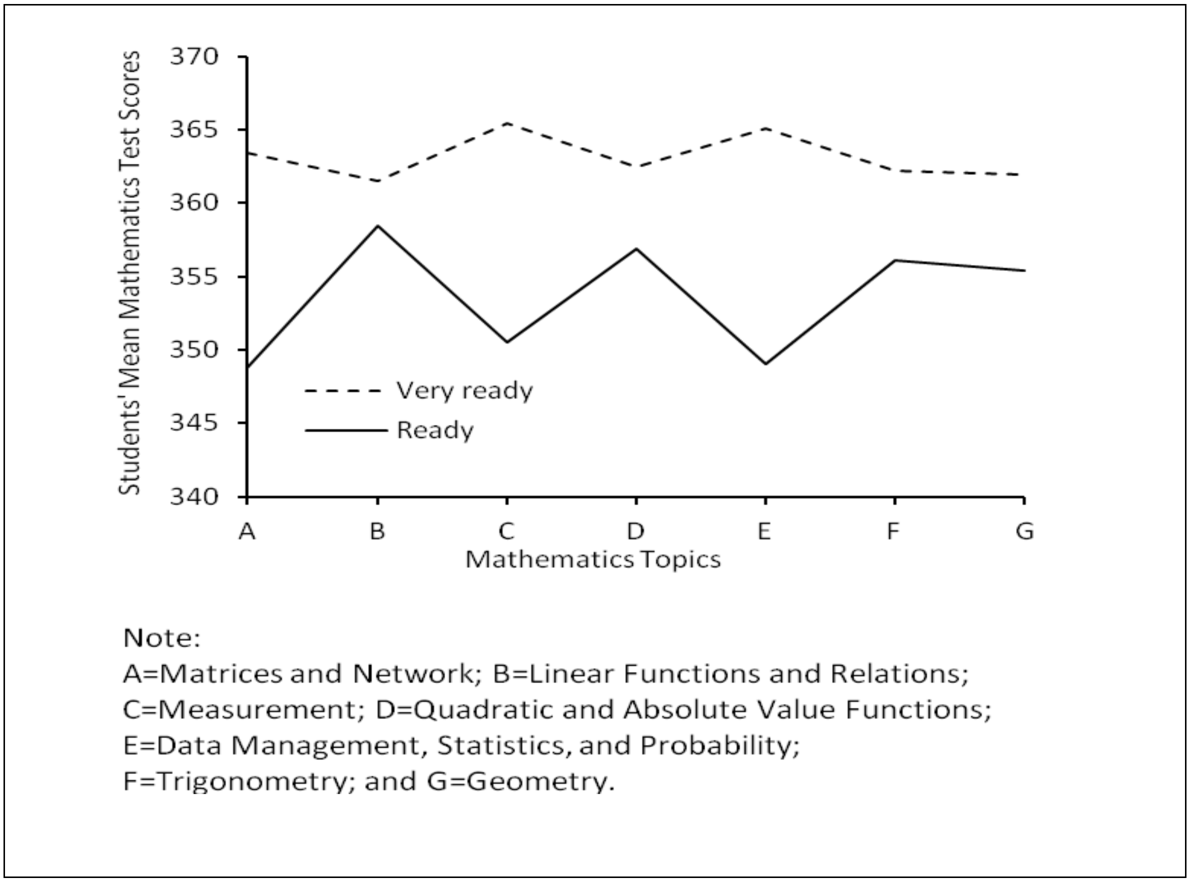


Figure 7. 3. Teachers' Readiness to Teach Grade 10 Mathematics Topics by Students' Mean Mathematics Test Scores

7.10.2 Teachers' Views about Their Ability to Teach and about their Job Satisfaction

Teachers were asked to report on what they thought about their ability to teach mathematics. Table 7.21 shows that over 85% of the students were taught by teachers who felt that they were capable of teaching mathematics to their students.

Table 7. 21. Percentages of Students Whose Teachers Reported Agreeing or Strongly Agreeing to Variables of Their Ability to Teach and Job Satisfaction

Measures	%	SE
I feel that I am making a significant educational difference in the lives of my students	96.1	9.81
If I try hard, I can make progress with even the most difficult and unmotivated students	90.3	9.95
I am successful with the students in my class	85.9	9.69
I usually know how to get through to students	92.3	8.66

Further, teachers were asked to rate their job satisfaction on a single item, “All in all, I am satisfied with my job”. The majority of students (83.4%, SE=9.60) were taught by the teachers who agreed or strongly agreed to the statement.

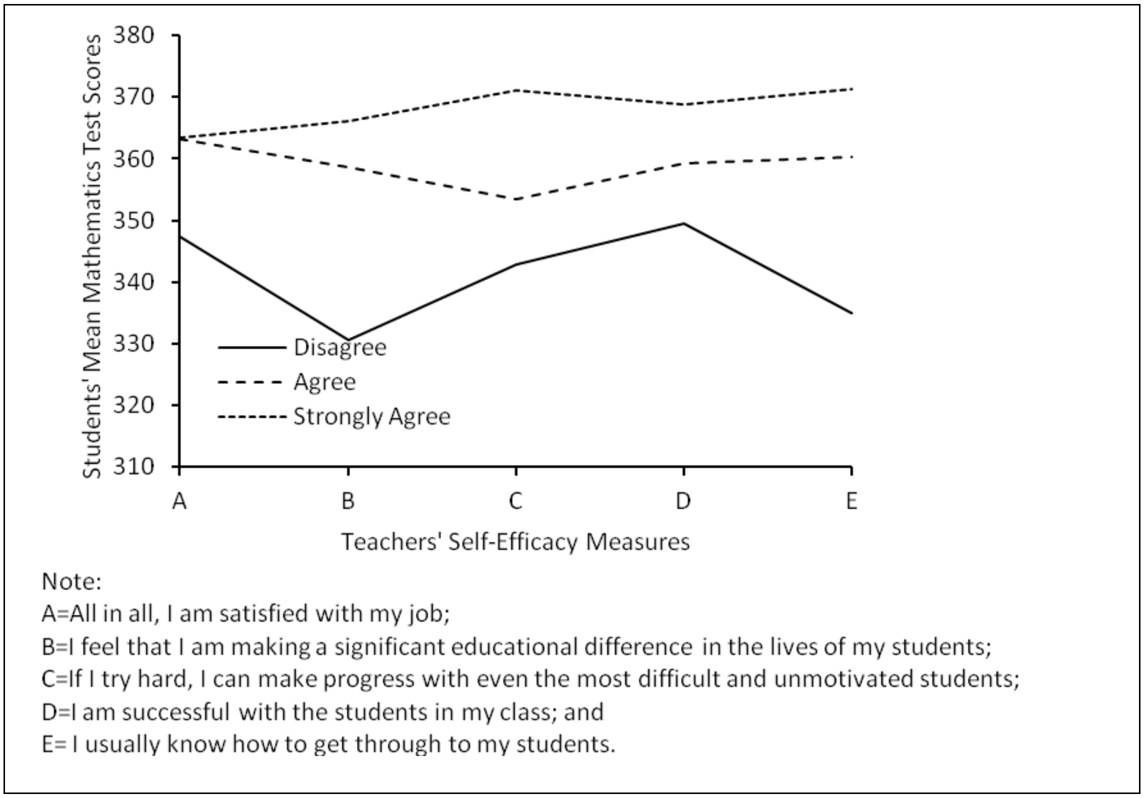


Figure 7. 4. Teachers' Self-Efficacy by Students' Mean Mathematics Test Scores

As depicted in Figure 7.4 (where line graph is used for the categorical data to emphasise the patterns which are not obvious in a column graph.), the students taught by teachers who strongly agreed or agreed to self-efficacy measures scored higher average mathematics test scores than the students taught by teachers who disagreed with the self-efficacy measures.

Discussion

The majority of students were taught by teachers who reported high self-efficacy and high readiness to teach the Grade 10 mathematics curriculum, which is an indication of the positive outlook of the teachers about their ability to teach mathematics. In addition, the finding shows that the teachers were aware of their potential as teachers. Similar to the findings reported in the literature (Caprara, et al., 2003; Goddard & Goddard, 2001; Goddard, et al., 2000; J. A. Ross, 1992; Tschannen-Moran, et al., 1998), the students who were taught by teachers who reported a higher self-efficacy outperformed the students who were taught by teachers whose responses indicated a lower self-efficacy. This finding shows the importance of high self-efficacy for the teachers, implying the need for schools to provide their teachers with environments and experiences conducive for developing high self-efficacy.

7.11 Homework

Homework, as related to teachers, was analysed in terms of the teachers' report on its frequency, duration, and purposes.

7.11.1 Teachers' Report on Homework Frequency and Duration

Teachers were asked to report about the frequency and duration of mathematics homework that they assigned to their students. Table 7.22 shows that the majority of students (76.7%) were taught by teachers who assigned homework for almost every lesson. Further, the majority of students were taught by teachers who assigned homework that required 15 to 30 minutes to complete.

Table 7. 22. Percentages of Students Whose Teachers Reported Various Homework Frequencies by Students' Mean Mathematics Test Scores

Measures	%	SE	Mathematics Test Scores	
			Mean	SE
Frequency				
Every or almost every lesson	76.7	6.10	365.43	4.64
About half of the lesson	19.5	5.76	344.01	7.18
Some lessons	3.7*	2.63	349.21	13.97
Duration				
Fewer than 15 minutes	13.8*	5.04	386.65	16.63
15-30 minutes	49.3	7.12	354.89	5.36
31-60 minutes	32.5	6.59	358.40	5.14
61-90 minutes	3.9*	2.71	351.13	10.72
More than 90 minutes	0.5*	0.09	422.11	7.71

Note: * the number of students of the teacher respondents was fewer than 50.

In addition, the students' mean Mathematics test scores corresponding to different aspects of homework are shown in Table 7.22. Overall, students' performance tended to increase when their teachers assigned them homework of 31-60 minutes duration for almost every lesson or every lesson.

7.11.2 Teachers' Use of Mathematics Homework

Teachers were asked how often they assigned different kinds of homework to their students. Table 7.23 shows that more than three-quarters of the students were taught by teachers who assigned homework on doing problems almost on a daily basis. However, fewer than one percent of the students were taught by teachers who reported assigning homework on gathering data and reporting on a daily basis.

Table 7. 23. Percentages of Students Whose Teachers Reported Assigning Different Types of Homework and Using Homework for Different Purposes Always or Almost Always and Sometimes

Measures	Always or Almost always		Sometimes	
	%	SE	%	SE
Homework Assignment				
Doing problem/question sets	75.2	6.42	24.8	6.42
Gathering data and reporting	0.9	0.13	78.8	5.78
Finding one or more applications of the content covered	23.5	6.22	62.4	6.98
Use of Homework				
Monitor whether or not the homework was completed	76.9	6.49	23.1	6.49
Correct assignments and then give feedback to students	50.8	7.12	49.2	7.12
Have students correct their own homework in class	12.3	4.67	70.7	6.33
Use the homework as the basis for class discussion	27.2	6.13	72.8	6.13
Use the homework to contribute towards students' grades or marks	53.5	7.06	44.4	7.00

Table 7.23 also shows the various uses that teachers made of their students' homework. More than 76% of the students were taught by teachers who used students' homework almost daily to monitor whether or not the homework was completed, while fewer than 13% of the students were taught by teachers who involved students in correcting their own homework. About a quarter of the teachers used students' homework as the basis for class discussion almost daily.

As shown in Figure 7.5, (where a bar graph has been used because of no response in some categories) teachers who always or almost always used their students' homework in different ways were associated with students who scored better than the students of teachers who used their students' homework sometimes or never.

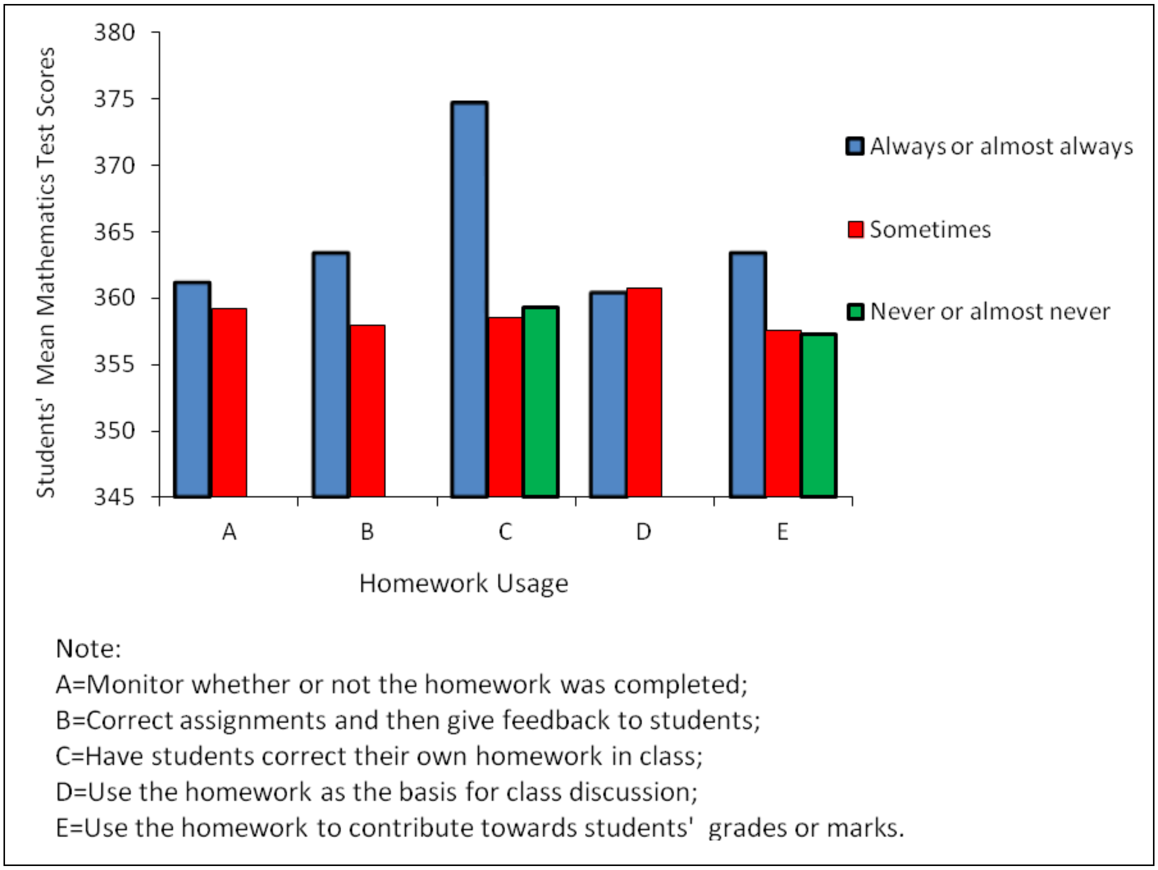


Figure 7. 5. Teachers' Usage of Homework by Students' Mean Mathematics Test Scores

Further, as shown in Figure 7.5, among the five different ways in which the teachers used their students' homework, having students correct their homework in the class corresponded to the highest test scores.

Discussion

The majority of teachers assigned homework of 15 to 30 minutes for every lesson and that most of the homework required students to solve mathematical problems. The analyses also revealed that the majority of teachers used students' homework for monitoring, giving feedback, group correction, class discussion, and awarding grades. These findings indicate a high prevalence of homework in schools and a range of uses that the teachers make of their students' homework in their teaching. While homework has a range of positive effects on student performance, too much of it is known to have adverse effects on student performance (Brock, et al., 2007; Cooper, et al., 2006). Similar to the findings reported in Cooper and Valentine (2001), student performance tended to peak when the students were assigned a homework of 31 to 60 minutes for almost every lesson. In addition, student performance tended to peak when teachers engaged students in correcting their homework in the class. The presence of a particular time range and a particular method of assessment as the determinants of the efficiency of homework imply that Bhutanese teachers need to plan how much homework they should give their students, what methods of assessment they should use to assess it, and what feedback strategies they should use for homework.

7.12 Tests and Examinations

Data on tests and examinations were analysed in terms of the types of mathematics tests and examinations used by the teachers. Data on the types of question formats used by teachers in tests and examinations are also analysed in this section.

7.12.1 Teachers' Use of Mathematics Tests or Examinations

Teachers were asked to provide information about how often they used mathematics tests or examinations in their class. Table 7.24 shows that almost two-thirds of the students were taught by teachers who gave a test in mathematics every month.

Table 7. 24. Percentages of Students Whose Teachers Reported Different Frequencies of Mathematics Tests or Examinations

Measures	%	SE	Mathematics Test Scores	
			Mean	SE
About once a week	1.4	0.20	335.04	4.10
A few times a week	16.6	5.22	356.85	8.96
About every two weeks	16.5	5.55	348.17	9.27
About once a month	65.5	6.82	365.31	5.18

As shown in Table 7.24, the students taught by the teachers who conducted mathematics tests about once a month performed at a higher level on the Mathematics test than the students taught by the teachers who conducted testing more frequently.

7.12.2 Teachers' Use of Question Formats in Tests

Teachers were asked to report on the types of questions they used in mathematics tests. As shown in Table 7.25, the large majority of the students, almost four-fifths, were taught by teachers who usually used questions involving application of mathematical procedures.

Table 7. 25. Percentage of Students Whose Teachers Reported Using Different Levels of Questions in Mathematics Tests Always or Almost Always and Sometimes

Questions	Always or Almost always		Sometimes	
	%	SE	%	SE
Questions involving application of mathematical procedures	77.3	5.92	22.7	5.92
Questions involving searching for patterns and relationships	20.3	6.23	77.1	6.40
Questions requiring explanations or justifications	19.2	5.18	80.8	5.18

Table 7.25 also shows that the teachers' use of questions on mathematical procedures was more frequent than their use of questions on mathematical concepts.

Discussion

Analyses of data on teachers' use of tests in mathematics classes showed that the majority of teachers administered mathematics tests once every month. In addition, the students taught by teachers who administered tests about once a month performed at a higher level than the students taught by teachers who administered tests more frequently. The analyses also revealed that most of the teachers used questions that demanded that students apply mathematical procedures, indicating a heavy emphasis on procedural mathematical knowledge and skills (OECD, 2010). Approximately, one-fifth of the teachers reported using questions that required students to identify patterns and relationships and formulate explanations and justifications to arrive at correct solutions, concurring with the OECD (2010).

7.13 Teachers' Usage of Textbooks

As Table 7.26 shows, the large majority of the students were taught by teachers who used textbooks as the primary basis of instruction. Fewer than 11% of the students were taught by teachers who used the textbook only as the supplementary resource for classroom instruction.

Table 7. 26. Percentages of Students Whose Teachers Reported Different Usage of Textbooks by Students' Mean Mathematics Test Scores

Textbook Usage	%	SE	Mathematics Test Scores	
			Mean	SE
Primary basis for lessons	89.1	4.40	361.23	4.46
Supplementary resource	10.9	4.40	356.96	4.24

Interestingly, the difference between the test scores of the students taught by teachers who used mathematics textbooks as the primary basis for lessons and the students taught by teachers who used mathematics textbooks only as a supplementary resource was statistically non-significant, *diff*=10.6, CI [-8.90, 30.16].

Discussion

Almost 90% of the teachers reported using textbooks as the primary basis for lessons as compared with fewer than 11% of the teachers who reported using the textbook as a supplementary resource for teaching. This finding emphasises the importance of the role played by textbooks in teaching and learning mathematics in Bhutanese schools, and also signals the severity of the consequences should Bhutanese schools be short-supplied with textbooks. To mitigate teachers' heavy reliance on textbooks if they come to be short supply, professional development on using alternative resources (e.g., the Internet, Library), should be provided to teachers.

7.14 Teachers' Use of Calculators

This section presents the analyses of data on the use of calculators in mathematics class as related to teachers. The analyses were done in terms of the availability of calculators in mathematics classes and the uses that the teachers made of the calculators in their classes.

7.14.1 Teachers' Report on the Availability of Calculators in Mathematics Classes

Teachers were asked how many students had access to calculators or graphing calculators. Table 7.27 shows that approximately 55% of the students were taught by teachers who reported having access to calculators. On the other hand, more than 81% of the students were taught by teachers who reported not having graphing calculators in mathematics classes.

Table 7. 27. Percentages of Students Whose Teachers Reported the Availability of Calculators in Mathematics Classes

Types of Calculators	All		Most		About half		Some		None	
	%	SE	%	SE	%	SE	%	SE	%	SE
Calculator	55.4	7.07	30.8	6.41	9.9	4.83	3.0	2.17	0.9	0.13
Graphing Calculator			10.9	5.35			7.6	3.69	81.5	6.11

7.14.2 Teachers' Use of Calculators in Mathematics Classes

Teachers were asked how they used calculators in their mathematics classes. Table 7.28 shows that over one-half of the students were taught by teachers who reported asking students to use calculators to check answers, to do routine computations, or to solve complex problems.

Table 7. 28. Percentages of Students Whose Teachers Reported Using Calculators for Various Purposes in the Mathematics Class

Measures	Every or Almost Every Lesson		About half the lessons		Some lessons		Never	
	%	SE	%	SE	%	SE	%	SE
Check answers	57.2	7.08	16.5	5.09	26.4	6.53		
Do routine computations	51.2	7.11	23.1	6.15	25.7	6.07		
Solve complex problems	51.4	7.11	22.0	5.93	21.8	5.78	4.9	2.83
Explore number concepts	18.4	5.78	21.2	5.75	41.8	6.98	18.5	5.51

Discussion

The analyses of data on students' access to, and usage of, calculators and computers in mathematics classes, as reported by teachers, revealed that one-half of the students had access to ordinary calculators in all classes and four-fifths of the students did not have access to graphing calculators. As reported in Ellington (2003), disparities in students' access to calculators may disadvantage the students without access to a calculator by slowing the growth in their abilities to solve mathematical problems and by fostering negative attitudes towards mathematics. This view is supported by the fact that the majority of teachers engaged their students in the use of calculators for checking answers, doing routine computations, and solving complex problem. Some of these situations might have caused learning difficulties for students who did not have access to calculators.

7.15 Teachers' Use of Computers in Mathematics Classes

Teachers were asked how often they used computers in their mathematics classes. Table 7.29 shows that more than almost two-thirds of the students were taught by teachers who never used computers in their classes.

Table 7. 29. Percentages of Students Whose Teachers Reported Using Computers for Various Purposes in Mathematics Classes

Measures	Every or Almost Every Lesson		About half the lessons		Some lessons		Never	
	%	SE	%	SE	%	SE	%	SE
Discover mathematics principles and concepts					23.9	6.28	76.1	6.28
Practise skills and procedures	5.2	3.83	1.8	1.33	13.1	4.73	80.0	5.83
Look up ideas and information			5.4	2.80	31.7	6.84	62.9	6.97
Process and analyse data			3.2	2.25	18.7	5.52	78.1	5.80

Discussion

While the use of calculators in mathematics classes is high, the use of computers in mathematics classes is very low. On average, approximately one in 20 teachers reported using computers in mathematics classes. Considering that computers have diverse potential to help students understand mathematical concepts, practise mathematical procedures, construct mathematical models, look up ideas and information, and analyse data (Ruthven, et al., 2004), not using computers in mathematics classes is a concern that will need to be addressed in the Bhutanese education system.

7.16 Summary and Implications for Policy

This chapter presented a range of key teacher characteristics that have potential to guide policy-makers in developing teacher-oriented educational interventions for improving school effectiveness. Overall, teachers' demographic profiles, professional collaborations, professional development, appraisals, views about classroom climate, pedagogical beliefs and practices, and self-efficacy correlated with student achievement. Further, teachers' usage of student homework, tests, and textbooks also correlated with student achievement.

Teachers' demographic profiles included gender, age, qualification, and teaching experience. The profiles showed that the majority of mathematics teachers in Bhutan were male, that the most effective teaching age of teachers ranged from 50 to 59 years, and that the teachers' qualifications and experience were correlates of student achievement. These findings suggest that more females be encouraged to take up mathematics during pre-service training, that age and experience of teachers be used as one of the dimensions of teacher deployment policies, and that teaching responsibilities of teachers be commensurate with their qualifications.

Teachers' professional collaborations comprised their discussion about mathematical concepts, their collaborative work on instructional materials, and their observation of one another's lessons. The majority of teachers reported discussing mathematical concepts with colleagues, and engaging in teamwork in preparing instructional materials. However, few teachers reported observing their colleagues' lessons or having their lessons observed by colleagues. Overall, teachers' professional collaborations did not relate well with student achievement; no interpretable pattern emerged from the univariate analysis of the index of professional collaborations and student achievement. These findings suggest that teachers' professional collaborations were not informed by the contemporary understanding of effective teaching. Such understanding might be developed by encouraging teachers to frame agendas for professional collaborations by observing one another's lessons, and to substantiate the collaborations with current research-based knowledge on effective teaching.

Teachers' professional development programmes need currency, relevance, and adequacy. Over one-half of the teachers reported not having participated in any professional development programmes in the past two years despite their need for the same. Little support from schools and fewer professional development programmes were some of the reasons cited by teachers for their inability to participate in professional development programmes. This suggests that professional development be informed by teachers' needs, and that teachers be provided with necessary support for their smooth participation in professional development programmes. The significance of professional development programmes for teachers was apparent in the finding that students taught by teachers with a low level of need for professional development programmes performed better on the Mathematics test than students taught by teachers with a high level of this need.

Teacher appraisal, as a platform for effective communication between teachers and school heads or between teachers and education monitoring officers, related to higher student achievement. However, not all teachers were aware of the level of importance attached to each appraisal criterion by their school heads or by education monitoring officers. These findings indicate a patchy quality of school leadership that needs improvement. This suggests that school heads or education monitoring officers need to communicate teacher appraisal criteria or teacher performance standards to teachers in advance of their appraisals.

Teachers' views about classroom climate provided interesting insights. First, the majority of students were taught by teachers who had favourable opinions about their school climate. Second, the majority of students were taught by teachers who had positive opinions about the classroom climate, with fewer than 20% of the students taught by teachers who encountered classroom disciplinary problems. However, teachers' views did not relate to a predictable pattern in student achievement on the Mathematics test. Teachers may have expressed false positive views on school and classroom climates.

Teachers' dispositions about teaching and learning strategies have the potential to help schools develop interventions to bring about effective teaching and learning. Overall, the percentage of teachers who believed in a constructivist teaching approach was greater than the percentage of teachers who believed in a transmissive teaching approach. However, the percentage of teachers who applied constructivist-teaching strategies was fewer than the percentage of teachers who applied transmissive teaching strategies. These findings suggest ambiguity in teachers' understanding of the relationship between teaching paradigms and teaching practices that may need to be clarified by providing more learning opportunities for teachers to enhance their repertoire of teaching strategies. Further, the majority of teachers reported using prescribed textbooks as the primary basis of instruction, suggesting the need for the teachers to be more resourceful in teaching and learning rather than being heavily reliant on prescribed textbooks. Implications of such interventions for student achievement are evident in the finding that students tended to perform better on the Mathematics test when their teachers reported having applied both the constructivist and the transmissive teaching approaches concurrently.

Teachers' capacity to teach effectively was also constrained by student characteristics and educational resources. The majority of teachers reported diverse student abilities and backgrounds, shortage of teaching and learning aids, and high student to teacher ratio as some of the major constraints. However, educational resources did not relate to student achievement in the Mathematics test, suggesting that the teaching strategies used by teachers who experienced fewer constraints were not different from the teaching strategies used by teachers who experienced more constraints. A policy implication from the findings is that relevant educational interventions be developed on the efficient use of teaching and learning resources.

Teachers' had a high readiness for, and a high self-efficacy about, teaching. The majority of teachers viewed themselves to be very ready for teaching. Similarly, over four-fifths of the teachers reported having a high self-efficacy about mathematics. Both readiness for teaching and self-efficacy about mathematics related well with student achievement in the Mathematics test. These findings suggest that teachers were knowledgeable in the contents of their teaching subject, and that teachers were provided with opportunities for developing their self-efficacy about mathematics.

The efficiency of homework was found to depend on its frequency, duration, and feedback. Homework was most efficient if it was assigned for almost every lesson. The effective time-on-task for a single homework occasion ranged from 31 to 60 minutes. Homework was also most efficient when teachers provided students with feedback on their work, and used it as a formative tool. These findings suggest that the frequency, the time-on-task, and the feedback aspects of homework be included in every school's homework policy.

Teachers reported using tests to assess the mathematical knowledge and skills of their students. The majority of teachers used such tests at least once a month; this frequency of testing was also the most efficient frequency because it corresponded to the highest mean Mathematics test score. The majority of teachers also reported using more procedural questions than conceptual questions in tests. These findings suggest that schools monitor the questions used by teachers in their tests, or teachers be trained in test development.

Calculators and computers were used by teachers in their mathematics classes for a range of purposes. The majority of teachers reported the use of calculators by their students for checking answers, doing routine computations, and solving complex problems in almost every lesson. Some policy implications from the findings are that students be encouraged to use calculators as tools for learning mathematical concepts and for solving complex mathematical problems. On the other hand, over two-thirds of students were taught by teachers who reported never having used computers in their mathematics classes. This indicates that teachers need to increase the use of ICT in their lessons, and that all classrooms be equipped, as soon as possible, with appropriate ICT hardware and software.

7.17 Chapter Conclusions

This chapter showed a wide range of teacher characteristics that have potential to improve the quality of the Bhutanese school education system. Teachers' demographic profiles, professional collaborations, professional developments, appraisal, classroom climate, school climate, educational resources, beliefs, teaching strategies, readiness for teaching, self-efficacy, homework, and assessment practices provided potential areas for educational interventions. Students taught by teachers who were strong in these characteristics performed better than students taught by teachers who were weak in these characteristics. This suggests a number of policy implications and strategic educational interventions. However, the potential of the teacher characteristics to improve the quality of the school education system in Bhutan will only be useful once the relevant educational interventions have found their way into Bhutanese schools.

Chapter 8

ANALYSES OF THE SCHOOL QUESTIONNAIRE

8.1 Introduction

The purposes of this chapter are to report the analyses of data from the school questionnaire and to identify key school characteristics that relate to student achievement. Where relevant, the findings from these data have been compared with the OECD and TIMSS's data to benchmark the Bhutanese education system against the education systems of other countries. Overall, the chapter addresses the research outcomes 2, 7, and 8 of generating knowledge about: factors related to effective schooling and their effects on student outcomes; educational structures and practices that maximise the learning opportunities; and equity in, and accessibility of, educational resources and provision. Specifically, the chapter answers the following research questions from Section 4.3.2.1:

ScL 1 How well do school policies and practices of student admittance, ability grouping, assessment and accountability, and parental involvement relate to student achievement? (Sections 8.3.1 through 8.3.5)

ScL 2 What is the state of school autonomy in Bhutan? (Section 8.3.6)

ScL 3 How well do school resources relate to student achievement? (Section 8.5)

ScL 4 How well do schools rate on the school climate scale? (Section 8.2)

ScL 5 How well do schools rate on an instructional leadership model? (Section 8.4)

ScL 6 What is the state of access and equity in school resources? (Section 8.5)

InL 1 What educational resources are available in schools? (Section 8.5)

The chapter has six sections: the first section reports on the factors of school climate, the second section reports on the school policies and practices, the third section reports on the instructional leadership, and the fourth section reports on the school resources. The fifth section presents policy implications—drawn on the basis of the patterns and statistics reported in the earlier sections—to signal the possible areas where policy directives can be formulated to guide relevant educational interventions. The last section concludes the chapter. The analyses were done in line with the analytical framework presented in Chapter 5.

8.2 School Climate

School principals were asked to provide their perceptions of a range of factors related to school climate. The factors were grouped into the following dimensions: student behaviour, teacher behaviour, teacher consensus, and teacher morale. The following sub-sections present the results from the analyses of data on the specific measures of these dimensions.

8.2.1 Student-Related Factors Affecting School Climate

Table 8.1 shows six student-related factors that affect school climate, and the percentages of students whose principals perceived the extent to which the factors hindered student learning. Table 8.1 also shows the OECD's average percentages which are adapted from the OECD (2004a, pp. 406-407).

Table 8. 1. Percentages of Students in Schools where Principals Reported Various Student-Related Variables Affected their School Climate to Some Extent or a Lot

Factors	Bhutan		OECD	
	%	SE	%	SE
Student absenteeism	48.7	7.63	48.4	0.6
Students skipping classes	34.1	6.96	30.3	0.6
Students using alcohol or illegal drugs	14.3	4.61	9.9	04
Disruption of classes by students	**9.9**	**3.59**	**40.0**	**0.6**
Students lacking respect for teachers	**9.0**	**3.44**	**22.0**	**0.6**
Students intimidating or bullying other students	**5.5**	**3.13**	**14.8**	**0.4**

Note: Bold numbers indicate that the difference between Bhutan's and the OECD's percentages are statistically significant at $p<0.05$.

As shown in Table 8.1, about one-half of the students were in schools whose principals identified student absenteeism as the most frequent student-related barrier to learning, followed by about one-third of the students being in schools whose principals identified students skipping classes as the next most frequent student-related obstacle to learning. Fewer than one-seventh of the students were in schools whose principals also reported use of alcohol or illegal drugs by students, disruption of classes by students, students' lack of respect for teachers, and students intimidating or bullying other students as barriers to learning.

Comparing Bhutanese and the OECD results, student-related factors that hindered teaching and learning were not as challenging in Bhutan as they were in the OECD countries, as shown by Table 8.1. For instance, disruption of classes by students, students lacking respect for teachers, and students intimidating or bullying other students were not as serious in Bhutan as they were in the OECD countries.

A linear regression analysis with the index of student-factors (STUDBEHA) as an independent variable and student performance on the Mathematics test as a dependent variable revealed that the amount of variance explained by STUDBEHA in the students' Mathematics test scores was statistically non-significant, R^2=0.002, 95% CI [-0.007, 0.01]. A unit increase in STUDBEHA corresponded to a statistically non-significant decrease of 1.00 score on the Mathematics test, 95% CI [-6.66, 3.24].

Discussion

Among the various student-related factors that affected the Bhutanese school climate, student absenteeism and truancy appear to be posing serious challenges to schools in delivering effective teaching and learning. For instance, almost one-half of the principals reported student absenteeism as the major problem followed by about one-third of the principals reporting students skipping classes as another major problem. The finding implies student disengagement with their schools (Finn & Voelkl, 1993). Student absenteeism and truancy also seem to be posing similar challenges in the OECD countries, indicating that they are global challenges to educators. However, Bhutanese schools appear to have fewer student-related disciplinary problems than the schools in the OECD countries. For example, the disciplinary problems caused by disruption of classes by students, students' lack of respect for teachers, and students intimidating or bullying other students in the schools in the OECD countries are statistically significantly greater than the disciplinary problems caused by the same factors in Bhutanese schools.

8.2.2 Teacher-Related Variables Affecting School Climate

Table 8.2 shows seven teacher-related variables affecting school climate and the corresponding percentages of the students whose principals perceived that the variables hindered student learning either to some extent or a lot. Table 8.2 also shows the OECD's average percentages as adapted from the OECD (2004a, pp. 411-412).

Table 8. 2. Percentages of Students in Schools where Principals Reported Teacher-Related Variables Affected their School Climate to Some Extent or a Lot

Factors	Bhutan		OECD	
	%	SE	%	SE
Teachers not meeting individual students' needs	41.6	7.64	33.3	0.6
Teachers' low expectations of students	**36.8**	**6.86**	**22.1**	**0.5**
Teacher absenteeism	25.0	6.13	18.9	0.6
Staff resisting change	18.8	5.54	25.7	0.5
Poor student-teacher relations	18.0	5.29	16.7	0.6

Students not being encouraged to achieve their full potential	17.1	5.12	23.2	0.6
Teachers being too strict with students	13.5	4.61	9.1	0.4

Note: Bold numbers indicate that the difference between Bhutan's and the OECD's percentages are statistically significant at $p<0.05$.

As shown in Table 8.2, teachers not meeting students' individual needs were the most frequent obstacle to student learning, with approximately two-fifths of the students enrolled in schools where principals perceived it as a hindrance to effective learning. The next most frequent obstacle was the teachers' low expectations of students: approximately one-third of the students were in schools whose principals perceived this as an obstacle to effective learning. This was followed by other hindrances such as teacher absenteeism, staff resisting change, poor-student-teacher relations, students not being encouraged to achieve their full potential, and teachers being too strict with students. However, the adverse effects of the teacher-related variables on student learning in Bhutan were not very different from those in the OECD countries. Only the teachers' low expectations of students were more prevalent in Bhutan than in the OECD countries, as shown in Table 8.2.

A linear regression analysis with the index of teacher-related variables (TEACBEHA) as an independent variable and student performance on the Mathematics test as a dependent variable showed that the amount of variance explained by TEACHBEHA in students' Mathematics test scores was statistically non-significant, R^2=0.002, 95% CI [0.01, -0.01,]. A unit increase in TEACBEHA corresponded to a statistically non-significant decrease of 2.27 scores on the Mathematics test, 95% CI [-8.37, 3.84].

Discussion

Amid other teacher-related variables, principals perceived the failure of their teachers to meet students' individual learning needs and the teachers' low expectations of their students as the two greatest impediments to student learning. Good (1987) noted that teachers' expectations of students may not necessarily be congruent with students' abilities, but rather they may be founded on teachers' inappropriate knowledge of how to respond to students' learning difficulties. This claim suggests that the teachers' inability to meet students' individual learning needs as the cause for the teachers' low expectations of their students. Further, the percentage of Bhutanese teachers holding low expectations of their students was significantly higher than the percentage of the teachers in the OECD countries, indicating the need for educational interventions aimed at encouraging Bhutanese teachers to have higher expectations of their students, and to inculcate such expectations in their students. Similar to the finding reported in the OECD (2004a), schools where principals reported more positive perceptions of the teacher-related factors affecting school climate seemed to perform better than schools where principals reported less positive perceptions. However, the relationship was not statistically significant.

8.2.3 Teacher Consensus

The prevalence of consensus among teachers in Bhutanese schools, as reported by school principals, is shown in Table 8.3.

Table 8. 3. Percentages of Students in Schools where Principals Agreed or Strongly Agreed with the Teacher Consensus Variables

Factors	%	SE
There are frequent disagreement between innovative and traditional mathematics teachers	51.7	7.74
There are frequent disagreements between mathematics teachers who consider each other to be ‘too demanding’ or ‘too lax’	32.5	6.89
There are frequent disagreements between mathematics teachers who consider each other as ‘too focussed on skill acquisition’ or ‘too focussed on the affective development of the student’	41.8	7.07

Table 8.3 shows that at least one-third of students were in schools where principals perceived frequent disagreements between teachers.

A linear regression analysis with the index of teacher consensus (TCCONS) as an independent variable and student performance on the Mathematics test as a dependent variable, however, showed that the amount of variance explained in the students’ mathematics test scores by TCCONS was statistically non-significant, R^2=0.001, 95% CI [-0.00, 0.01]. A unit increase in TCCONS corresponded to a statistically non-significant decrease of 1.21 scores on the Mathematics test, 95% CI [-4.32, 1.20].

Discussion

Teacher consensus, as a factor affecting school climate, appeared to be an issue in schools. More than one-third of the students were in schools whose principals reported that their teachers disagreed on their views about teaching approaches, expectations for students, and lesson priorities. Such disagreements among teachers may hinder professional collaborations, peer teaching, lesson observation, and transfer of knowledge and skills between teachers, resulting in the loss of opportunities to improve teaching practices (Hargreaves, 2001). This finding indicates the needs for schools to develop interventions capable of fostering consensus among teachers on their views of teaching approaches, expectations for students, and lesson priorities.

8.2.4 Teacher Morale and Commitment

Table 8.4 shows the percentages of the students in schools where principals either strongly agreed or agreed to four aspects of teacher morale. Table 8.4 also shows the OECD's average percentages of students across the OECD countries where principals reported on the four aspects of teacher morale (OECD, 2004a, p. 413).

Table 8. 4. Percentages of Students in Schools where Principals Agreed or Strongly Agreed with the Teacher Morale Variables

Factors	Bhutan		OECD	
	%	SE	%	SE
Teachers take pride in this school	100.0	9.33	90.0	0.4
Teachers value academic achievement	100.0	9.42	93.1	0.2
The morale of teachers in this school is high	**34.5**	**6.87**	**87.2**	**0.4**
Teachers work with enthusiasm	**33.1**	**6.83**	**89.7**	**0.4**

Note: Bold numbers indicate that the difference between Bhutan's and the OECD's percentages are statistically significant at $p<0.05$.

As shown in Table 8.4, all the students were enrolled in schools where principals either agreed or strongly agreed to the statement: "teachers take pride in this schools" and "teachers value academic achievement". On the other hand, only one-third of the students were in schools where principals perceived their teachers as having high morale. Similarly, approximately one-third of the students were in the schools where principals perceived their teachers being enthusiastic with their work. In addition, Table 8.4 shows that the percentages of students in schools where principals' perceptions teachers' morale and work enthusiasm in Bhutan were significantly lower than the corresponding OECD averages.

A linear regression analysis with the index of teacher morale and commitment (TCMORALE) as an independent variable and student performance on the Mathematics test as a dependent variable showed that the amount of variance explained in the students' Mathematics test scores by TCMORALE was statistically non-significant, R^2=0.007, 95% CI [-0.10, 0.02]. A unit increase in TCMORALE corresponded to a statistically non-significant decrease of 2.32 scores on the Mathematics test, 95% CI [-5.49, 0.85].

Discussion

Teacher morale and commitment, as perceived by school principals, gave further insights into school climate. On the one hand, all the principals reported that their teachers took pride in their schools and valued academic achievement, indicating teachers' affinity with their schools and desire for their students to excel in studies. On the other hand, approximately one-third of the principals reported that their teachers did not have high morale and that their teachers did not work with enthusiasm. Although this finding is hard to explain, it supports the prevailing public perceptions of teachers in the Bhutanese education system. Literature has linked teacher morale to factors such as supportive leadership, appraisal and recognition, curriculum coordination, effective discipline policy, excessive work demands, goal congruence, participative decision-making, professional growth, professional interaction, role clarity, and student orientation (Hart, Wearing, Conn, Carter, & Dingle, 2000; Lumsden, 1998; Mackenzie, 2007). The connection of teacher morale to these factors show the broader issues that may need to be addressed due to low teacher morale on the overall school climate, indicating an urgent need for interventions to boost teacher morale.

8.3 School Policies and Practices

School principals were asked to provide perceptions of their schools' policies and practices related to the following areas: student admittance, student and teacher assessment, school accountability, student ability grouping, parental engagement, and school autonomy. The following sub-sections present the results of the analyses of data on these areas.

8.3.1 School Admission Policies

As shown in Table 8.5, diverse criteria were set by schools for admitting students. As indicated in Table 8.5, similar student admission policies were also followed across the OECD countries (OECD, 2007b, p. 161), but to varying extents.

Table 8. 5. Percentages of Students in Schools where Principals Reported School Admission Criteria as Pre-requisites or as High Priorities

Criteria	Bhutan		OECD	
	%	SE	%	SE
Residence in a particular area	69.0	8.97	49.6	0.5
Directives from the Ministry of Education*	53.8	8.24	--	--
Students' academic record	39.4	7.86	26.7	0.5
Recommendation of feeder schools	32.6	7.28	12.6	0.3
Students' need or desire for a special programme	17.4	5.71	18.9	0.5
Attendance of other family members	13.6	5.09	16.5	0.4

Note: * not applicable to OECD

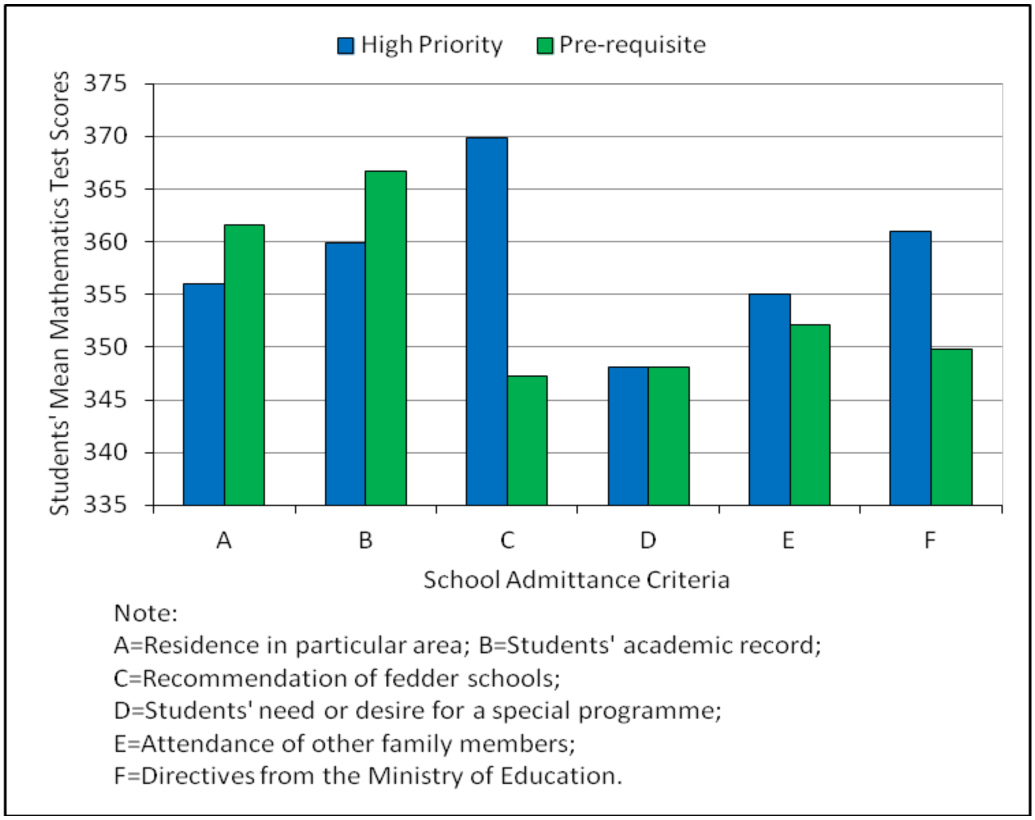


Figure 8. 1. School Admittance Criteria by Students' Mean Mathematics Test Scores

The relationship between the severity of the school admittance criteria and students' mean Mathematics test scores, depicted in Figure 8.1, shows that except for the recommendation of feeder schools and the directives from the Ministry of Education, severity in school admittance criteria did not relate strongly to student achievement on the Mathematics test.

Discussion

The dominance of students' place of residence and the Ministry's directives as priorities in school admissions policies imply the existence of such policies in schools. This finding also indicates fewer opportunities for students to choose or change schools and fewer options for schools to admit high-performing students, which may result in less competition among schools for student enrolment and less parental pressure on the schools to perform better (OECD, 2007a; Soderstrom & Uusitalo, 2010). Given that an open school admission policy is known to enhance student achievement, student and teacher relationships, and student satisfaction at school (Cullen, et al., 2005; Gibbons, et al., 2008; Lavy, 2010), students' residence and the Ministry's directives as dominant priorities may need to be relaxed to allow greater student mobility. However, the lack of consistency in prioritising the various elements of admission policies by schools implies policy differences among schools and the extent of freedom for some schools to develop such policies, indicating the needs for the Ministry of Education to standardise school admission policies.

8.3.2 Assessment Policies and Practices

School principals provided information about a range of frequencies with which different assessments were conducted in their schools. As shown in Table 8.6, the majority of students were in schools where a range of assessments was practised either twice a year or at least three times a year. Table 8.6 also shows the OECD's average percentages based on data from the OECD (2004a, p. 420). Similar patterns were present in assessment practices in Bhutan and the OECD countries, but with variation in frequencies.

Table 8. 6. Percentage of Students in Schools where Principals Reported Conducting the Following Assessments either Two Times a Year or Fewer and Three Times a Year or More in their Schools

Assessment	Two times a year or fewer				At least three times a year or more			
	Bhutan		OECD		Bhutan		OECD	
	%	SE	%	SE	%	SE	%	SE
Standardised tests	53.5	7.97	77.0	0.6	46.5	8.32	23.0	0.6
Teacher-developed test	19.6	5.15	8.5	0.3	46.9	8.00	91.5	0.3
Teachers' judgmental rating	47.7	8.73	25.3	0.5	52.3	8.75	74.7	0.5
Student portfolio	54.1	7.90	56.7	0.6	45.9	8.54	43.3	0.6
Student assignment/project work/homework	15.2	4.69	14.1	0.4	58.9	9.44	85.9	0.4

The analysis of the relation between the frequency of different assessments and students' mean Mathematics test scores shows an interesting relationship as in Figure 8.2.

Overall, the test scores of the students in schools where principals reported using assessments at least three times in an academic year tended to be higher than the test scores of the students in schools where principals reported using assessments twice or fewer in an academic year.

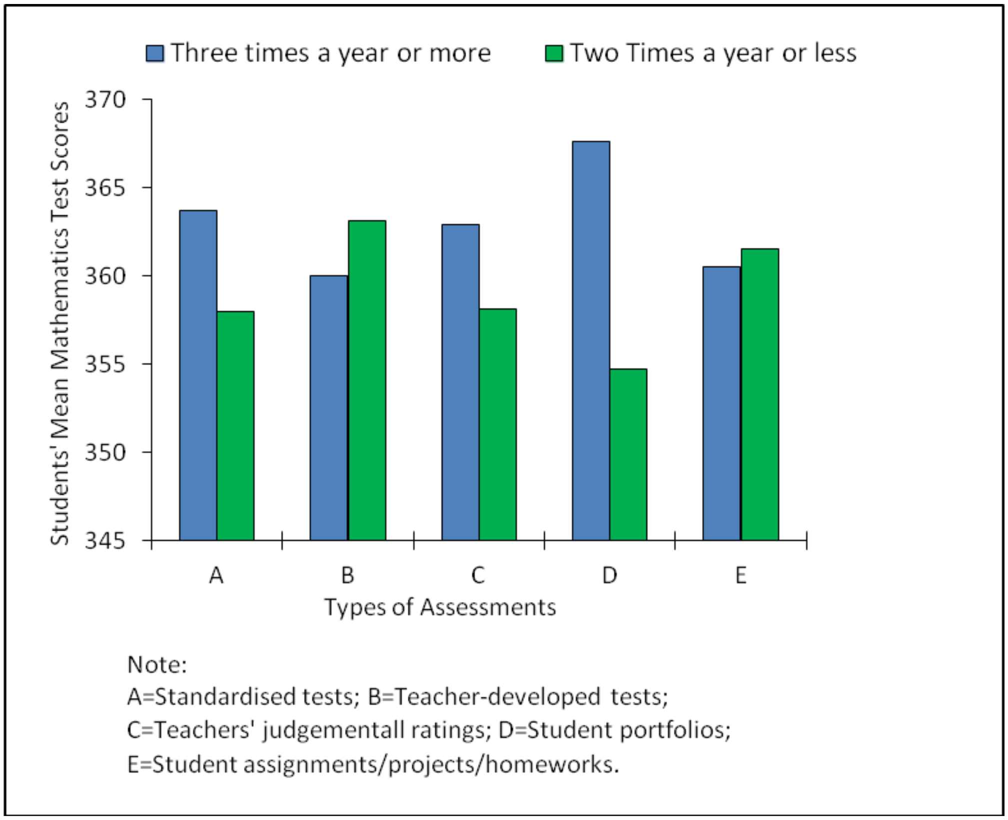


Figure 8. 2. Types of Assessments by Students' Mean Mathematics Test Scores

Discussion

Schools practised a range of assessments: standardised tests; teacher-developed tests; teachers' judgemental ratings; student portfolios; and student assignments. This finding is indicative of the rich assessment knowledge and skills that are available in the schools and the existence of relevant policies that guide the use of the assessments. Furthermore, the association between the frequent use of assessments and student achievement is in line with the findings reported in the OECD (2004a).

8.3.3 Assessment and Accountability

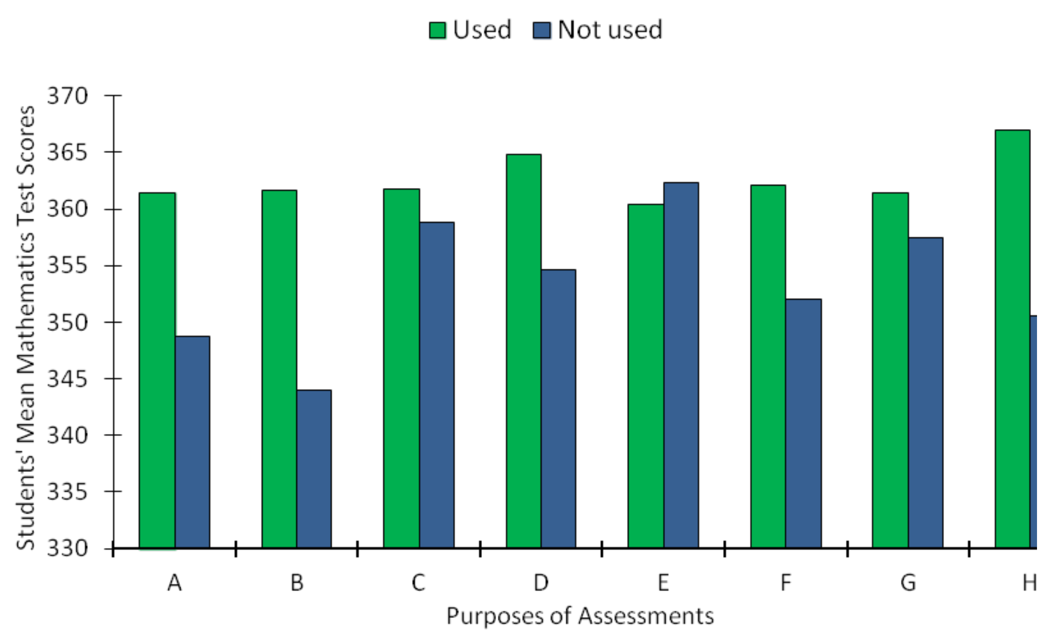
The majority of students were in schools where principals reported using assessments for different purposes, as shown in Table 8.7. Table 8.7 also shows OECD average percentages as reported by the OECD (2004a, pp. 421-424). There were significant differences in the use of the school assessments in Bhutan and in the OECD countries as instruments of accountability. The difference was most apparent in the use of assessments to make judgements about teachers' effectiveness. This was followed by the use of assessments to compare schools, and monitoring the schools' progress from year to year.

Table 8. 7. Percentages of Students in Schools where Principals Reported Different Uses of Assessment Results

Assessment Purpose	Bhutan		OECD	
	%	SE	%	SE
To make decisions about students' retention or promotion	**94.3**	**3.24**	**78.9**	**0.4**
To inform parents about their child's progress	93.4	3.37	95.1	0.3
To monitor the school's progress from year to year	**89.8**	**4.59**	**69.3**	**0.5**
To make judgments about teachers' effectiveness	**85.0**	**4.84**	**43.9**	**0.6**
To identify aspects of instruction or the curriculum that could be improved	79.7	5.31	74.3	0.5
To compare the school with other schools	**61.4**	**6.79**	**40.4**	**0.6**
To group students for instructional purposes	**61.1**	**6.80**	**43.4**	**0.6**
To compare the school to national performance	59.0	6.98	45.8	0.5

Note: Bold numbers indicate that the difference between Bhutan's and the OECD's percentages are statistically significant at p<0.05.

Figure 8.3 shows the relationship between the students mean Mathematics test scores and the ways in which the school principals reported using school assessments.



Note:
A=To inform parents about their child's progress;
B=To make decisions about students' retention or promotion;
C=To group students for instructional purposes;
D=To compare the school to national performance;
E=To monitor the school's progress from year to year;
F=To make judgements about teachers' effectiveness;
G=To identify aspects of instruction or the curriculum that could be improved;
H=To compare the school with other schools.

Figure 8. 3. Purposes of Assessments by Students' Mean Mathematics Test Scores

The test scores of students in schools where school principals reported using school assessments for various purposes, as shown in Figure 8.3, were consistently higher than the test scores of students where principals reported not using the assessments for the same purposes.

Further, school principals were asked if they used student achievements, teacher peer review, and observation of lessons for monitoring the classroom practices of mathematics teachers at their schools. Table 8.8 shows that the majority of students were in schools where principals reported using tests, teacher peer review, and lesson observations by principals or senior teachers for monitoring teacher performance. On the other hand, fewer than one-fifth of the students were in schools where principals reported using observation of classes by inspectors or persons external to the school to monitor their teachers. As shown in Table 8.8, similar practices of monitoring school teachers were used across the OECD countries (OECD, 2004a, pp. 432-433); though, the scope of the usage was much larger in Bhutan than in the OECD countries.

Table 8. 8. Percentages of Students in Schools where Principals Reported Prevalence of Different Teacher Monitoring Tools in their schools

Teacher Monitoring Tools	Bhutan		OECD	
	%	SE	%	SE
Tests or assessments of student achievement	**76.7**	**5.81**	**58.5**	**0.7**
Teacher peer review	**80.6**	**5.29**	**53.7**	**0.7**
Observation of lessons by principals or senior teacher	**95.9**	**2.41**	**60.7**	**0.5**
Observation of classes by inspectors or other persons external to the school	28.9	6.60	24.5	0.6

Note: Bold numbers indicate that the difference between Bhutan's and the OECD's percentages are statistically significant at $p<0.05$.

Further, the relationship between the approaches to monitoring teachers and student achievement was analysed. Figure 8.4 shows the results of the analyses.

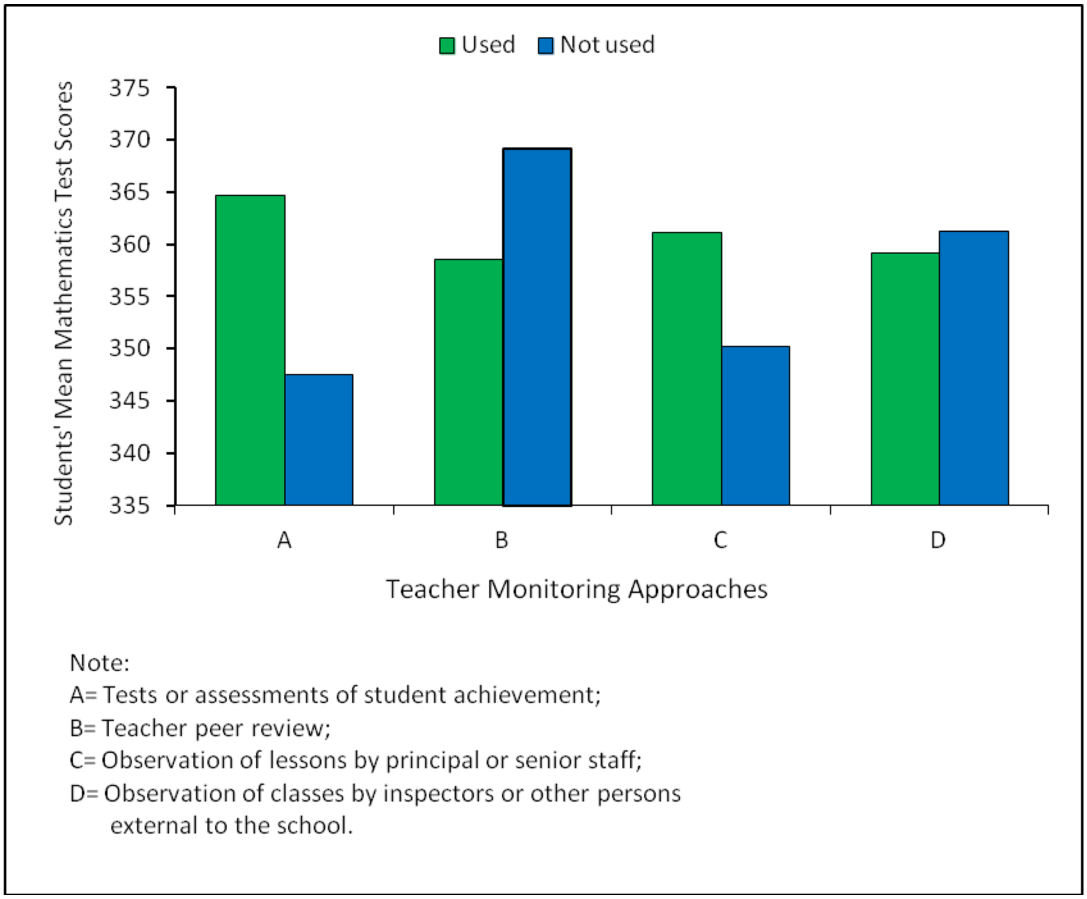


Figure 8. 4. Teacher Monitoring Approaches by Students' Mean Mathematics Test Scores

As shown in Figure 8.4, the principals' use of tests or assessments of student achievement had resulted in maximum difference in students' test scores. The next most influential teacher monitoring approach was the observation of teachers' lessons by principals or senior staff as shown by the difference in the test scores of the students whose principals reported using the teacher monitoring approach and not using the approach. The teacher peer review as an approach to monitoring teachers' practices did not result in a positive influence on the students' test scores; rather, the use of teacher peer review was associated with lower test scores as compared to the test scores associated with not using it.

Discussion

Principals reported using assessments for accountability on different fronts. On the student front, student assessments were used for deciding student promotion, informing parents of their child's progress, and grouping students for instructional purposes. On the teacher front, student assessments were used for judging teacher effectiveness and identifying instructional strategies that could be improved. On the school front, student assessments were used for monitoring school's progress from year to year, comparing schools with other schools, and benchmarking schools' performance with the national averages. The use of assessments for accountability purposes in Bhutanese schools was significantly greater than that in the schools in the OECD countries (OECD, 2004a). In addition, the schools' use of assessments for maintaining accountability related to student achievement, with the students in schools that used the assessments for accountability purposes doing better than the students in schools that did not use the assessments for such purposes.

8.3.4 Ability Grouping within School

School principals were asked if their mathematics teachers used ability grouping—for all classes or for some classes or not for any classes—as an approach to cater to students' learning needs. Eight percent (SE=4.28) of the students were enrolled in schools where principals reported that ability grouping was used by the teachers for all classes. Fifty-four percent (SE=7.07) of the students were in schools where principals reported that their teachers used ability grouping for some classes. Thirty-eight percent (SE=6.77) of the students were in schools where principals reported that ability grouping was not used for any classes in their schools.

The analyses of schools' practice of ability grouping and student achievement showed that the former did not contribute to the latter. In other words, the difference between the mean Mathematics test scores of the students whose principals reported their teachers using ability grouping for all classes and not using ability grouping for any classes was statistically non-significant, *diff*=1.33, 95% CI [-30.42, 33.09]. Similarly, the use of ability grouping for some classes and not for any classes did not result in a statistically significant difference between the test scores of the corresponding groups of students, *diff*=0.28, 95% CI [-16.39, 15.83].

Discussion

The practice of ability grouping in schools was at best variable, and did not relate to student performance. The percentages of principals who reported that their teachers used ability grouping ranged from as low as eight percent to as high as 50 percent, indicating the lack of uniformity in the practice of ability grouping by the schools. Similar to the findings reported in Betts and Shkolnik (2000) and the OECD (2004a), ability grouping did not lead to a difference in student achievement in Bhutanese schools.

8.3.5 Parental Engagement in School

School principals were asked if their schools expected students' parents to engage in some activities of their schools. Table 8.9 shows that the majority of students were in schools where principals expected the parents to participate in various school activities. Table 8.9 also shows the TIMSS average percentages as reported by Mullis et al. (2008, p. 341) for Grade 8 students. As shown in the table, schools' expectations of the parents to participate in school activities in Bhutan were very similar to schools' expectations of the parents in the countries that participated in TIMSS.

Table 8. 9. Percentages of Students in Schools where Principals Reported Expecting Parental Engagement in Various School Activities

Parents' Role	Bhutan		TIMSS	
	%	SE	%	SE
Ensure that their child completes his/her homework	98.0	1.54	95.0	0.3
Attend special events	97.0	3.00	90.0	0.4
Serve on school committees	74.7	6.27	71.0	0.5
Volunteer for school projects, programmes, and trips	74.5	6.41	84.0	0.5
Raise funds for the school	56.1	7.11	54.0	0.6

The analyses of the association between schools' expectations of the parents' engagement in school activities and student achievement gave useful insights as shown in Figure 8.5.

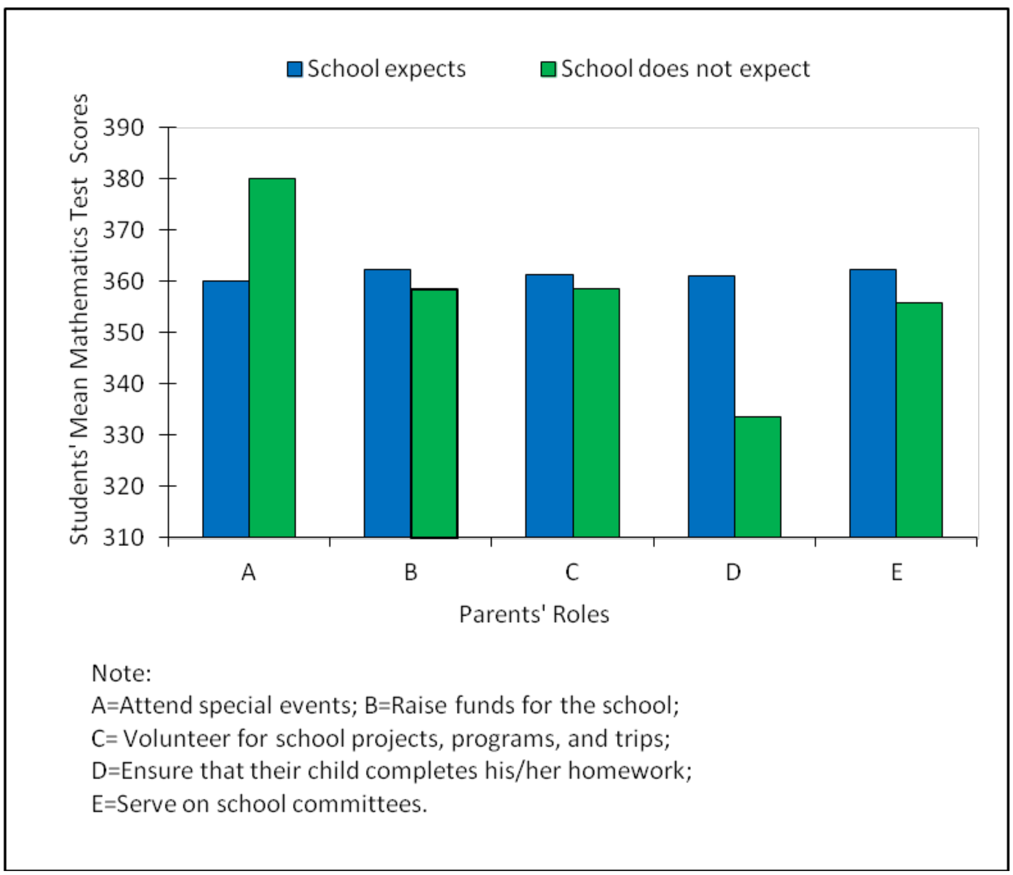


Figure 8. 5. Parents' Roles by Students' Mean Mathematics Test Scores

Overall, students in schools where principals reported that their schools encouraged the parents to engage in school activities achieved higher average Mathematics test scores than the students in schools where principals reported that their schools did not encourage the parents to engage in school activities. Notably, the average Mathematics test scores of the students in the schools where principals reported encouraging the parents to ensure that their child completed his/her homework was statistically significantly higher than the performance scores of the students in the schools where principals reported no such encouragement, *diff*=27.67, 95% CI [12.06, 43.28].

Discussion

The expectations of the school principals for the parents to be involved in school activities indicate the schools' desire to reach out to the parents. Similar to the findings reported in other studies (Bowen & Lee, 2006; Driessen, et al., 2005; Fan & Chen, 2001; Grolnick, et al., 1997; Grolnick & Slowiaczek, 1994; Izzo, et al., 1999; Jeynes, 2003, 2005, 2007; Simon, 2001, 2004; Yan & Lin, 2005), parental involvement in school activities is related to higher average Mathematics test scores in Bhutanese schools. The positive effect was very prominent where principals expected the parents to encourage their children to complete school homework.

8.3.6 School Autonomy

School principals were asked to identify stakeholders who shouldered the following responsibilities: formulating school budgets, deciding on budget allocations within the school, establishing school disciplinary policies, establishing student assessment policies, approving students for admittance to school, and choosing which textbooks were used. The percentages of the students in schools whose principals identified the stakeholders who shouldered these responsibilities are shown in Figure 8.5 (SMB is an abbreviation for School Management Board).

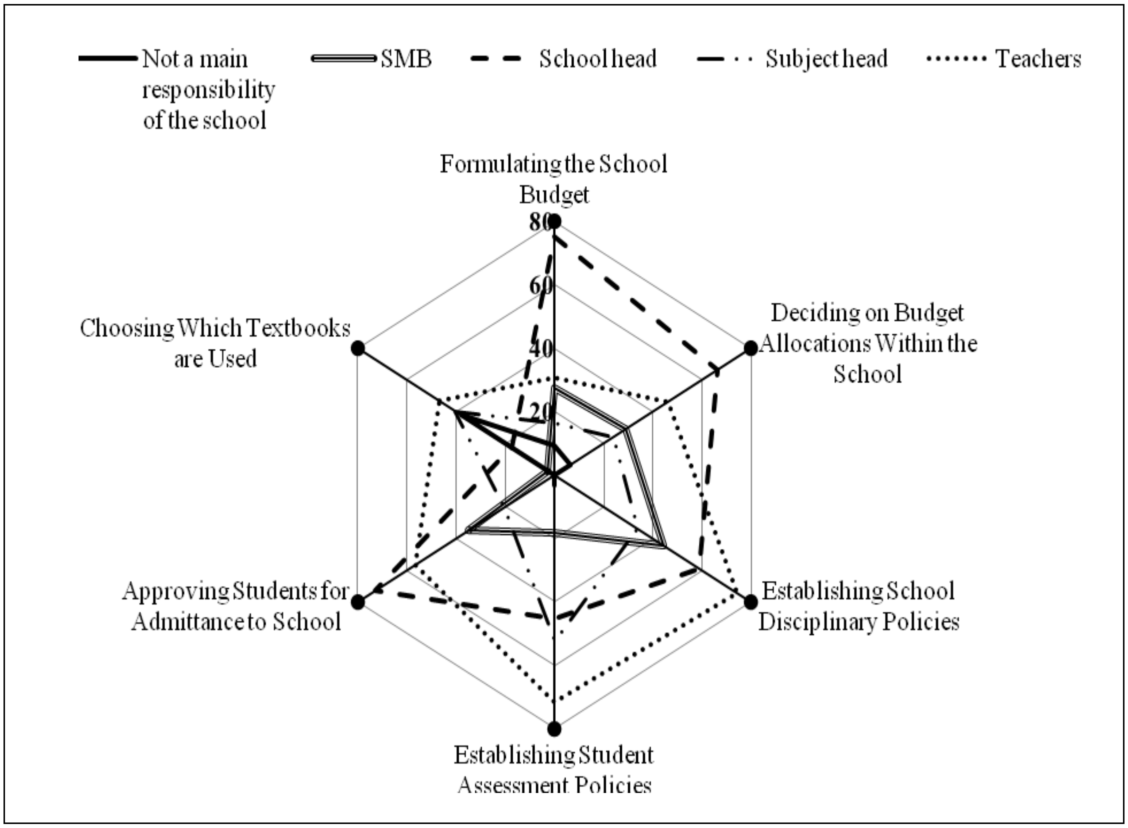


Figure 8. 6. Percentages of Students in Schools where Principals Reported the Responsibilities of Various Stakeholders in their Schools

As shown in Figure 8.6, the principals differed widely in their perceptions about the degree to which stakeholders should shoulder various responsibilities. For instance, 75% of the students were enrolled in schools where principals perceived that the formulation of the school budget was their responsibility. On the other hand, 28%, 16%, and 31% of the students were enrolled in schools where principals perceived that the formulation of the school budget was the responsibility of SMB, Subject Heads, and teachers respectively. Nine percent of the students were also enrolled in the schools where principals perceived that the formulation of the school budget was not their responsibility.

While information about stakeholders who shouldered various responsibilities indicate some degree of school autonomy, information about stakeholders involved in making decisions on the responsibilities provide further information about the extent of school autonomy. School principals were asked to identify stakeholders that they thought made decisions on the following areas: staffing, budgeting, instructional content, assessment practices. Figure 8.7 shows the extent to which stakeholders exert direct influence on decision making at schools. As shown in Figure 8.7, the majority of students were in schools where principals reported that the schools' budgeting and staffing were decided by the Dzongkhags [local administrative districts, with some responsibility for educational decision-making] or the Ministry of Education. Further, Figure 8.7 shows that the majority of students were in schools where principals reported that the parents had little influence on the decisions related to staffing, budgeting, instructional contents, and assessment practices of their children's schools.

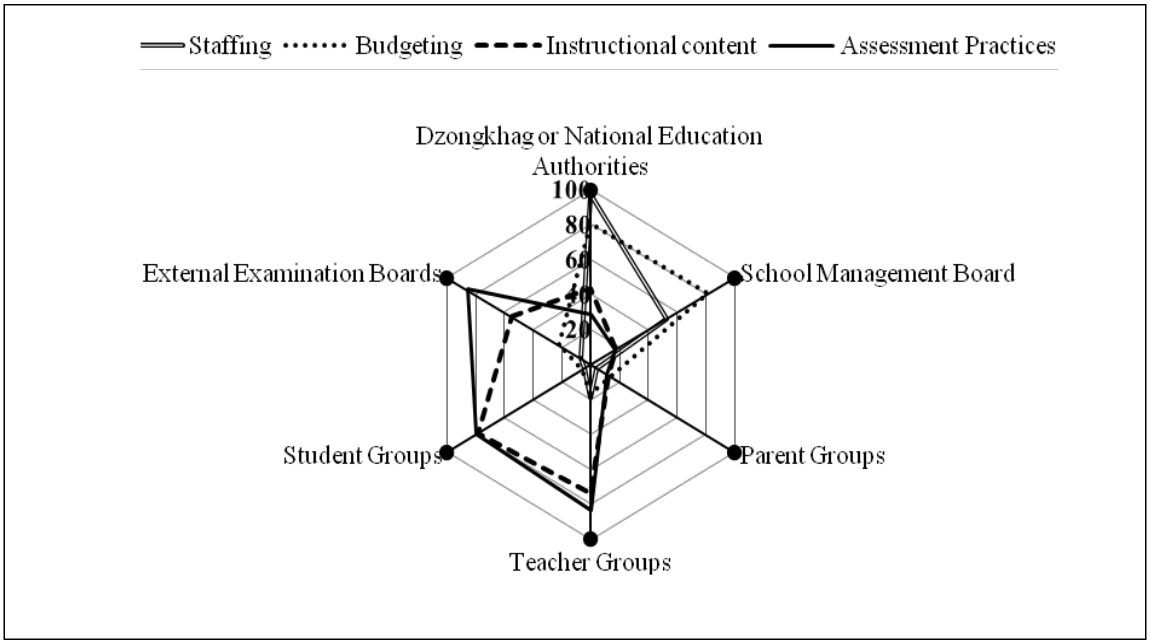


Figure 8. 7. Percentages of Students in Schools where Principals Reported Various Stakeholders and their Influences in Different Areas in which Schools Make Important Decisions

Discussion

School principals differed widely in their views of school autonomy, as characterised by the responsibilities of formulating school budget, deciding on budget allocation within the school, establishing school disciplinary policies, establishing school assessment policies, approving students for admittance to school, and choosing textbooks. This indicates the lack of uniformity among the principals in their understanding of their responsibilities. This finding may also be indicative of a communication gap between the Ministry of Education and the schools. However, schools were almost consistent in their report of stakeholders who made decisions on staffing, budgeting, instructional contents, and assessment practices. As expected, decisions on staffing and budgeting were made by the Ministry of Education or the school districts. Overall, the inconsistencies in school principals' views about who should take what responsibilities, and who should make the decisions, indicates a need for better communication between the stakeholders because school autonomy has been clearly linked to better school performance (Fuchs & Wobmann, 2007; Maslowski, et al., 2007; OECD, 2010; West, et al., 2010; Wobmann, 2007).

8.4 School Leadership

School principals were asked to report on how frequently they performed their leadership roles, with the roles being: setting goals; communicating goals; supervising and evaluating instructions; coordinating curriculum; monitoring student progress; protecting instructional time; maintaining visibility; providing teacher incentives; providing professional development; and providing incentives for learning. The frequencies were categorised as follows: almost always, frequently, sometimes, seldom, and almost never. The following sections present the result of the analyses on these data for each of these leadership aspects.

Linear regression analyses were undertaken for each aspect of the principal's instructional leadership as the independent variables and student performance scores on the Mathematics test as the dependent variable resulted in statistically non-significant regression coefficients. However, the descriptive analyses of these data on the various aspects of instructional leadership revealed some patterns that have the potential to guide policy interventions related to improving school effectiveness, as shown in Table 8.10.

Table 8. 10. Percentages of Students in Schools where Principals Reported Performing Instructional Leadership Roles Frequently or Almost Always

Roles	%	SE
1. Goal Setting		
Develop a focussed set of annual school wide goals	82.5	9.57
Frame the school's goals in terms of staff responsibilities for meeting them	90.7	9.74
Use needs assessment or other formal and informal methods to secure staff input on goal development	76.7	8.12
Use data on student performance when developing the school's academic goals	62.0	9.25
Develop goals that are easily understood and used by teachers in the school	86.7	9.95
2. Communicating the Goal		
Communicated the school's mission effectively to members of the school community	80.9	9.84
Discuss the school school's academic goals with teachers at staff meetings	93.5	7.73
Refer to the school's academic goals when making curricular decisions with teachers	84.3	9.41
Ensure that the school's academic goals are reflected in highly visible displays in the school	79.6	9.48

Refer to the school's goals or mission in forums with students	82.0	9.44
3. Supervising and Evaluating Instruction		
Ensure that the classroom priorities of the teachers are consistent with the goals and direction of the school	87.6	9.94
Review student work products when evaluating classroom instruction	69.9	9.19
Conduct informal observations in classrooms on a regular basis	49.8	8.13
Point out specific strengths in teacher's instructional practices in post-observation feedback	61.8	8.14
Point out specific weaknesses in teacher instructional practices in post-observation feedback	55.4	7.96
4. Coordinating School Curriculum		
Make clear who is responsible for coordinating the curriculum across grade levels	91.3	9.92
Draw upon the results of school-wide testing when making curricular decisions	68.6	8.54
Monitor the classroom curriculum to see that it covers the school's curriculum objective	69.5	8.98
Assess the overlap between the school's curricular objectives and the school's achievement tests	39.3	7.42
Participate actively in the review of curriculum materials	57.8	8.63
5. Monitoring Student Progress		
Meet individually with teachers to discuss student progress	61.9	9.31
Discuss academic performance results with the faculty to identify curricular strengths and weaknesses	77.4	9.36
Use tests and other performance measures to assess progress toward school goals	69.2	9.36
Inform teachers of the school's performance results in written form	63.5	9.38
Inform students of the school's academic progress	88.6	10.03
6. Protecting Instructional Time		
Limit interruptions of instructional time by public address announcements	60.7	9.20
Ensure that students are not called to the office during instructional time	63.4	9.19
Ensure that tardy and truant students suffer specific consequences for missing instructional time	46.5	8.08
Encourage teachers to use instructional time for teaching and practising new skills and concepts	86.0	9.70
Limit the intrusion of extra-and co-curricular activities on instructional time	89.6	9.81
7. Maintaining High Visibility		
Take time to talk informally with students and teachers during recess and breaks	79.1	9.89
Visit classrooms to discuss school issues with teachers and students	67.8	9.43
Attend/participate in extra-and co-curricular activities	96.7	8.63
Cover classes for teachers until a late or substitute teacher arrives	67.5	8.87
Tutor students or provide direct instruction to classes	68.1	9.40
8. Providing Teacher Incentives		
Reinforce superior performance by teachers in staff meeting, newsletters, and /or memos	72.3	9.29
Compliment teachers privately for their efforts or performance	80.3	9.84

Acknowledge teachers’ exceptional performance by writing memos for their personnel files	53.5	8.16
Reward special efforts by teachers with opportunities for professional recognition	70.4	9.55
Create professional growth opportunities for teachers as a reward for special contributions to the school	71.9	9.68
9. Providing Professional Development		
Ensure that in-service activities attended by staff are consistent with the school’s goals	77.8	9.51
Actively support the use in the classroom of skills acquired during in-service training	80.2	9.77
Obtain the participation of the whole staff in important in-service activities	84.3	9.67
Lead or attend teacher in-service activities concerned with instruction	64.9	9.10
Set aside time at staff meetings for teachers to share ideas or information from in-service activities	84.2	10.00
10. Providing Incentives for Learning		
Recognize students who do superior work with formal rewards such as an honour roll or mention in the principal’s newsletter	80.9	9.76
Use assemblies to honour students for academic accomplishments or for behaviour or citizenship	96.9	9.69
Recognize superior student achievement or improvement by seeing in the office the students with their work	68.8	9.40
Contact parents to communicate improved or exemplary student performance or contributions	62.1	8.97
Support teacher actively in their recognition and/or reward of student contributions to and accomplishments in class	81.2	9.73

Table 8.10 indicated that the principals varied in performing the 10 leadership roles, with some leadership roles performed more frequently than the others. Table 8.11 shows the difference between the mean index of any one leadership role and the mean indices of the remaining leadership roles; the highest mean index indicates the most performed role. Table 8.11 is read from left to right. For instance, the Leadership Role 1 (goal setting) is compared to the Leadership Role 2 (communicating the goal), and shows that the Leadership Role 2 is performed more frequently than the Leadership Role 1.

Table 8. 11. Comparison of the Mean Scores of the Indices of Different Leadership Roles

Role			1	2	3	4	5	6	7	8	9	10
	Mean		1.22	2.01	0.70	0.50	0.83	0.55	1.15	1.05	1.40	1.24
		SE	0.23	0.24	0.25	0.19	0.21	0.15	0.20	0.22	0.27	0.22
1	1.22	0.23	0.00	**-0.79**	**0.52**	**0.72**	**0.39**	**0.67**	0.07	0.17	-0.18	-0.02
2	2.01	0.24	**0.79**	0.00	**1.31**	**1.50**	**1.18**	**1.45**	**0.85**	**0.96**	**0.61**	**0.77**
3	.70	0.25	**-0.52**	**-1.31**	0.00	**0.20**	-0.13	0.15	**-0.45**	**-0.35**	**-0.70**	**-0.54**
4	.50	0.19	**-0.72**	**-1.50**	**-0.20**	0.00	**-0.33**	-0.05	**-0.65**	**-0.55**	**-0.89**	**-0.73**
5	.83	0.21	**-0.39**	**-1.18**	0.13	**0.33**	0.00	**0.28**	**-0.32**	**-0.22**	**-0.57**	**-0.41**
6	.55	0.15	**-0.67**	**-1.45**	-0.15	0.05	**-0.28**	0.00	**-0.60**	**-0.50**	**-0.85**	**-0.69**
7	1.15	0.20	-0.07	**-0.85**	**0.45**	**0.65**	**0.32**	**0.60**	0.00	0.10	**-0.24**	-0.08
8	1.05	0.22	-0.17	**-0.96**	**0.35**	**0.55**	**0.22**	**0.50**	-0.10	0.00	**-0.35**	-0.19
9	1.40	0.27	0.18	**-0.61**	**0.70**	**0.89**	**0.57**	**0.85**	**0.24**	**0.35**	0.00	0.16
10	1.24	0.22	0.02	**-0.77**	**0.54**	**0.73**	**0.41**	**0.69**	0.08	**0.19**	-0.16	0.00

Note:

1. Bold numbers indicate statistically significant difference at alpha<0.05.
2. 1 represents goal setting; 2 represents communicating the goal; 3 represents supervising and evaluating instructions; 4 represents coordinating school curriculum; 5 represents monitoring student progress; 6 protecting instructional time; 7 maintaining high visibility; 8 represents providing teacher incentives; 9 represents providing professional development; and 10 represents providing incentives for learning.

Table 8.11 shows some unexpected differences between the frequencies with which the principals reported performing the leadership roles. For instance, the mean index column shows that the principals set school goals more frequently than they supervised instructions, coordinated school curriculum, protected instructional time, maintained high visibility, provided teacher incentives, or provided incentives for learning. Therefore, the role performance patterns in Table 8.11 indicate areas where policy interventions may be needed.

Discussion

School principals practised all aspects of instructional leadership. However, the principals seemed to be either confused with different aspects of instructional leadership roles or that they were not as cognizant of the roles as they might be. For instance, the mean index of goal setting was greater than the mean index of other leadership, as discussed above. This shows that the principals set goals more frequently than performing any other leadership roles, differing from the usual trend of setting school goals at the beginning of an academic year. Alternatively, the principals seemed to change school plans more often than implementing them. The principals' report of practising all the roles of instructional leadership, but not in expected patterns, suggests the need for the Ministry of Education to provide opportunities for the principals to enhance their instructional leadership knowledge and skills. Such a need is quite pressing, bearing in mind claims that the success of instructional leadership depends on the proficiency of the principals in its concept and associated implementation strategies (Hallinger & Heck, 1996, 1998; Maeyer, et al., 2007; Marks & Printy, 2003; Witziers, et al., 2003). The list of Roles in Table 8.10 might prove to be a valuable resource in planning in-service activities for principals designed to enhance their instructional leadership skills.

8.5 School Resources

School principals provided information about the state of their schools' human and material resources. The following sections present the analyses of data on these resources. The association between school resources and student learning was not statistically significant; therefore, only the percentages of students enrolled in the schools where principals perceived that the lack of resources hindered student learning is reported in the following sub-sections. Notwithstanding the lack of association between the student learning and the resources, the percentages reveal the state of school resources in the schools across Bhutan, as perceived by the principals.

8.5.1 Human Resources

School principals were asked to provide information about the extent to which the shortage of qualified and experienced mathematics teachers hindered the schools' capacity to provide instruction. School principals also provided information about the hindrances to instruction caused by the shortage of replacement teachers and support personnel. Table 8.12 shows the percentage of students in schools where principals reported the hindrances caused to the instruction by the lack of human resources.

Table 8. 12. Percentages of Students in Schools where Principals Reported that their Schools' Capacity to Provide Instruction was Hindered to Some Extent or a Lot by the Shortage or Inadequacy of Different Variables of Human Resources

Human Resource Shortage or Inadequacy	Bhutan	
	%	SE
Qualified mathematics teachers	51.4	8.25
Experienced mathematics teachers	58.2	8.01
Replacement teachers	64.2	9.00
Support personnel	38.7	7.89

As shown in Table 8.12, over 50% of students were in schools where principals perceived that the lack of qualified teachers or experienced teachers or replacement teachers hindered their schools' capacity to provide quality instruction. On the other hand, fewer than two-fifths of the students were in schools where principals reported that their schools' capacity to provide effective instruction was hindered by the lack of support personnel. Overall, Table 8.12 shows that schools have difficulty in accessing these resources, and by implication, different schools may have different levels of difficulty, which further suggests inequitable distribution of the resources.

Figure 8.8 shows the relationship between human resource constraints and students' mean Mathematics test scores. Except for experienced mathematics teachers and replacement teachers, students of the schools where principals reported that the schools' capacity to provide effective instruction was hindered by lack of qualified mathematics teachers and support personnel performed better than other students.

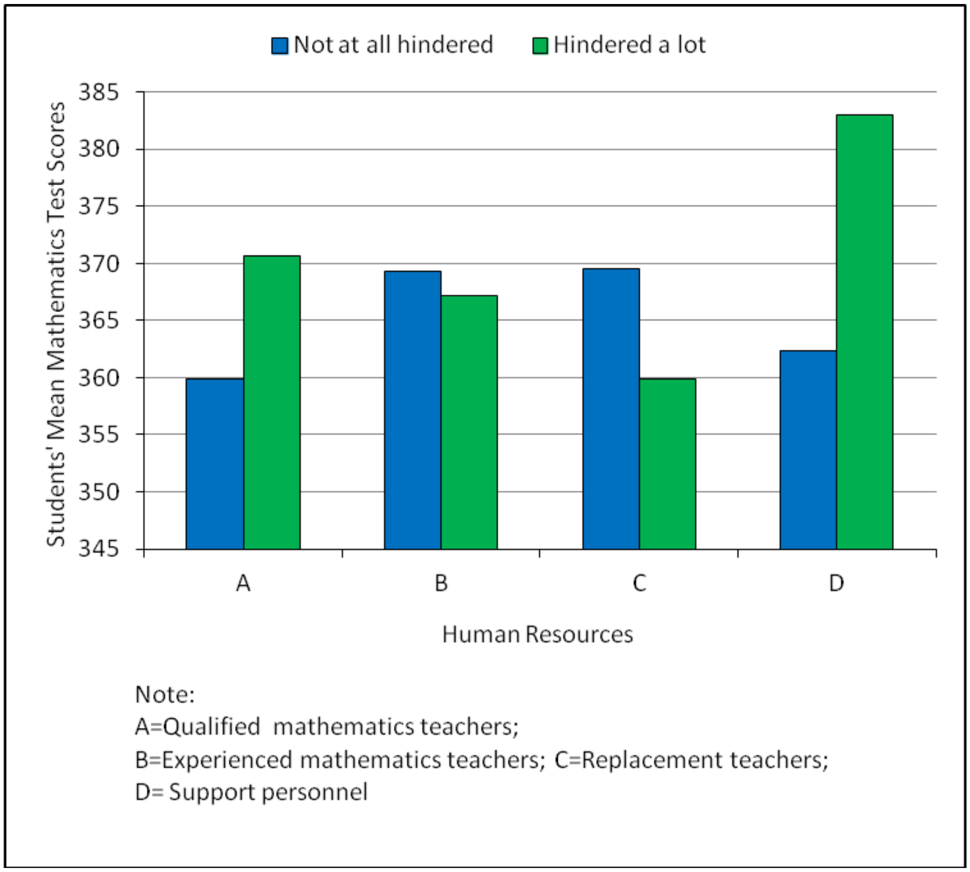


Figure 8. 8. Human Resources by Students' Mean Mathematics Test Scores

As shown in Figure 8.8, except for experienced mathematics teachers and replacement teachers, students of the schools where principals reported that the schools' capacity to provide effective instruction was hindered by lack of qualified mathematics teachers and support personnel performed better than other students.

Discussion

The majority of principals perceived that the problem posed by the lack of adequate human resources was serious. More than half the principals reported the shortage or inadequacy of qualified mathematics teachers, experienced mathematics teachers, and replacement teachers as hindering schools' capacity to provide effective instruction. However, the relationship between hindrances caused by the shortage or inadequacy of human resources and students' mean Mathematics test scores did not show a clear pattern in support of this statement, suggesting a gap between the principals' perceptions and practical advantages or disadvantages of the resources to student achievement that needs further study. Where principals' perceptions of the challenges presented by the shortage of human resources related to student outcomes suggests a need for more qualified and experienced teachers because "the quality of an education system cannot exceed the quality of its teachers" (Barber & Mourshed, 2007, p. 16). Further, as asserted by Scheerens, Glass, and Thomas (2003) that school resources are malleable, the Ministry of Education could provide more opportunities for its current teachers to upgrade their subject knowledge and pedagogical skills, and set higher standards for recruiting prospective teachers.

8.5.2 Material Resources

School principals were asked to provide information about the extent to which the shortage of instructional materials, computer software, calculators, library books, audio-visual resources, and laboratory equipment hindered the schools' capacity to provide effective instruction. Table 8.13 shows the percentages of students in schools where principals reported hindrances caused to the instruction by the lack of material resources.

Table 8. 13. Percentages of Students in Schools where Principals Reported that their Schools' Capacity to Provide Instruction was Hindered to Some Extent or a Lot by the Shortage or Inadequacy of the Variables of Material Resources

Material Resource Shortage or Inadequacy	Bhutan	
	%	SE
Instructional materials	59.0	7.71
Computers for instruction	64.2	8.13
Computer software for instruction	66.5	9.00
Calculators for instruction	49.2	7.80
Library materials	48.1	7.65
Audio-visual resources	65.8	9.06
Science laboratory equipment and materials	58.3	8.75

As shown in Table 8.13, more than one-half of the students were in schools where principals reported that the shortage or inadequacy of instructional materials, computer software for instruction, audio-visual resources, computers for instruction, and laboratory equipment and materials hindered the schools' capacity to provide quality classroom instruction. Percentages in all seven categories are very high.

Figure 8.9 shows the relationship between material resources and students' mean mathematics test scores. Similar to human resources, there is no discernible pattern in the relationship, indicating a gap between the principals' perspectives and student achievement.

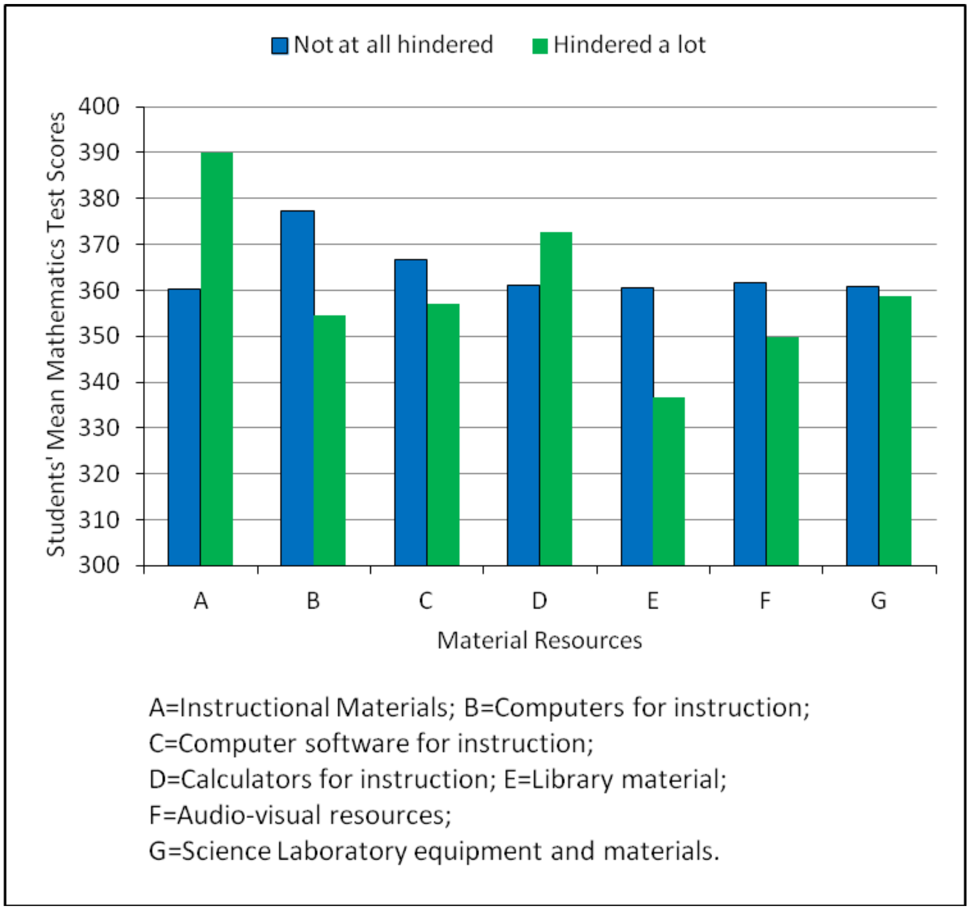


Figure 8. 9. Material Resources and Students' Mean Mathematics Test Scores

Discussion

Similar to human resources, about one-half of the principals reported that a shortage or inadequacy of material resources hindered schools' capacity to provide effective instructions. This finding shows that the schools were either under-resourced or that they were not adept at optimising the use of their resources. As emphasised in the literature on school resources, it is often the problem of not using the resources effectively rather than the problem of not having the resources (Greenwald, et al., 1996; Hanushek, 1997; Hedges, et al., 1994; Marzano, 2001; Scheerens, 2000; Wayne & Youngs, 2003; Wenglinsky, 2002). Further studies should be conducted into the schools' use of their resources, as well as programmes to enhance their provision.

8.6 Summary and Implications for Policy

The chapter presented several key school characteristics that were associated with student learning. School climate, school policies and practices, school leadership, and school resources were shown as correlates of student learning. Overall, the results from the analyses suggested that it is possible for Bhutan to improve its school education system by formulating relevant policies, with the aim of nurturing school characteristics that are conducive to improved student learning. This view is elaborated in the following paragraphs.

School climate consists of a multi-dimensional construct, with the dimensions being student-related factors, teacher-related factors, teacher consensus, and teacher morale. On the student-related factors, student absenteeism, and students skipping classes were reported by principals as the most frequent hindrances to student learning. The findings indicate the need for policy directives aimed at reducing student absenteeism, and student truancy. On the teacher-related factors, teachers not meeting individual students' needs, teachers' low expectations of students, and teacher absenteeism were reported by principals as the most frequent hindrances to student learning. In addition, staff resisting change, poor student-teacher relations, and students not being encouraged to achieve their full potential were reported as less frequent, but still noteworthy, hindrances to student learning.

The teacher-related factors, as a dimension of school climate, showed that teachers needed to cultivate greater consensus and nurture positive morale. School principals reported that approximately half of their mathematics teachers disagreed on teaching practices, teaching goals, and work expectations and attitudes. This finding shows a need for the school principals to facilitate frequent professional collaborations among teachers, with the aim of fostering consensus among them; especially, on the matters related to teaching and learning. On teacher morale, all the principals reported that their teachers took pride in their schools and that their teachers valued academic achievement. However, the majority of principals reported that their teachers were not enthusiastic about their work and that their teachers had low morale. This finding signals an urgent need for a policy intervention aimed at motivating teachers and boosting their morale as well as further studies into why teachers have low morale.

School policies and practices comprised school admission policies, assessment policies and practices, assessment and accountability, ability grouping within school, parental engagement in school, and school autonomy. First, the analysis of data on the school admittance policies showed that the majority of principals considered students' residence and directives from the Ministry of Education as priorities for admitting students to their schools. These were followed by other students' academic records and recommendations of feeder schools. The fact that the principals differed in using school admission policies underlines the lack of standard school admission policies across the schools in Bhutan, indicating a need for standardisation of school admission policies.

Assessment policies and practices, as a component of school policies and practices, showed that schools differed in the types of assessments they practised and the frequencies with which they practised the assessments. Students in schools where principals reported using assessments for three times or more in a school year performed better than students in schools where principals reported using assessments twice or less in the school year. The findings indicate that while schools used a range of assessments to evaluate different aspects of student learning, the frequency of assessments needs to be monitored. Schools should develop assessment policy guidelines to monitor the type and the frequency of assessments practised by their teachers so that the practices result in improved student learning.

Assessment and accountability, as a component of school policies and practices, revealed that assessment was a popular means of delivering accountability among schools. The majority of school principals reported using assessments for the following purposes: to make decisions about students' retention or promotion, inform parents about their child's progress, monitor schools' progress, make judgements about teachers' effectiveness, improve instruction and curriculum, compare schools or schools with national performance, and group students for instructional purposes. In addition, the analysis of data on assessments and accountability showed that students in schools where principals reported using assessments to establish accountability, on average, performed better on the Mathematics test than students in schools where principals reported not using data to establish accountability. These findings indicate that schools should be provided with policy guidelines on the use of assessments as tools of school accountability.

Ability grouping within schools, as a component of school policies and practices, showed that it was not well established in schools. Almost one-half of the students were in schools where principals reported that they did not practise it. Although ability grouping did not lead to a significant difference in student performance, its use as one of a range of student-oriented teaching strategies are perhaps worth pursuing in schools.

Parental engagement, as a component of school policies and practices, needs to be promoted in schools. The majority of principals reported expecting parents to attend special events, serve on school committees, volunteer for school projects, raise funds for the schools, and help children to complete homework. Students in schools where principals reported such expectations, on average, performed better on the Mathematics test than students in schools where principals did not expect parents to involve themselves in the school activities. These findings suggest a need for policy guidelines for engaging parents in school activities.

School autonomy, as a component of school policies and practices, showed that schools needed succinct guidelines on their areas of autonomy. The majority of principals reported having autonomy in formulating their school budget, deciding on budget allocation within their school, school disciplinary policies, and approving students for admittance to school. On the other hand, more than one-half of the principals reported a lack of autonomy in choosing textbooks for use in their schools, with about two-fifths of the principals considering it as not their responsibility. The school management board also played a significant role in establishing school disciplinary policies, approving students for admittance to school, and deciding on budget allocation within school. The findings indicate the lack of uniformity in the extent to which schools perceived autonomy as defined by the factors mentioned earlier. In addition, school principals reported that external examination boards, teacher groups, and student groups made decisions on the types of assessments practised in their schools. All principals reported that decision on staffing or teacher deployment was made by either the Dzongkhag education officers or by the Ministry of Education. Similarly, four-fifths of the principals reported that decisions on school budgeting were made by either the Dzongkhag or the Ministry of Education. The findings indicate little autonomy for schools in making decisions on staffing and budgeting, and this may require a policy change.

School leadership comprised ten dimensions as detailed in Table 8.10. Overall, the majority of principals reported practising all the ten aspects of the instructional leadership. The frequencies with which the principals performed the ten aspects of the instructional leadership were similar, though the frequencies were not as expected. For instance, principals reported setting goals more often than performing other leadership roles, though the goals were usually set once at the beginning of the school year. The findings suggest that the principals either performed the leadership aspects randomly throughout the school year or they were not well-acquainted with the instructional leadership model, suggesting another potential area for change.

School resources included human and material resources. On the human resource aspect of the schools, the majority of the principals reported that the shortage or inadequacy of qualified mathematics teachers, experienced mathematics teachers, and replacement teachers hindered schools' capacity to provide proper instruction. On material resources, the majority of the principals reported that the shortage or inadequacy of instructional materials, computers for instruction, computer software for instruction, calculators for instruction, library materials, audio-visual resources, and science laboratory equipment and materials hindered schools' capacity to provide effective instruction. These findings suggest that the Ministry of Education conduct further studies into school resources with the aim of formulating policy guidelines on procurement and enlarged distribution of resources to schools in terms of quantity, quality, and time.

8.7 Chapter Conclusions

This chapter identified a range of school characteristics that have potential to improve the Bhutanese school education system. The analyses in the chapter showed that there is scope for the Bhutanese education system to develop a positive school climate, standardise school policies and practices, enhance the calibre of school leadership, and improve efficiency in procurement and use of school resources. Further, drawing on the empirical evidence accrued from the analyses in this chapter, several major policy implications were suggested. These signal important areas where policy makers and other relevant stakeholders may consider enacting policies and introduce educational interventions to improve the Bhutanese education system.

Chapter 9

ANALYSES OF THE FOCUS GROUP INTERVIEW

9.1 Introduction

The purposes of this chapter are to report the analyses of data from the focus group interview and to identify emerging themes on the context-level educational effectiveness factors of the proposed national educational assessment model.

The purposes, the method, the participants, and the focus group size were fully documented in Chapter 4. They are also briefly recalled in this chapter for clarity. The chapter then presents the results of the critical analyses of the responses to the focus group interview questions. The questions and answers are grouped under each of the three context-level factors, analysed individually for emerging themes, and discussed. The chapter then presents a section each for policy implications and conclusions.

9.2 Stage for the Focus Group Interview

The participants for the focus group interview reported to the pre-arranged venue where the researcher and moderator greeted them. As the participants were familiar with each other, the moderator introduced the themes of the focus group interview. The introduction included the purpose of the focus group interview presented in Chapter 4. Briefly recalled, the purpose of the focus group interview was to explore the perspectives of the participants about the following three context-level educational effectiveness factors of the proposed national educational assessment model: external achievement stimuli (CoL 2); national education policies on effective learning environments (CoL 1); and mechanisms for evaluating national education policies (CoL 3). The introduction was followed by a few pleasantries between the participants and the moderator to relax the former for the focus group interview questions. The participants were then familiarised with some ground rules to ensure a collegial environment during the interview.

9.3 Critical Analyses of the Focus Group Interview Questions

The focus group interview questions and answers are analysed and the results are presented as sub-sections of the individual context-level educational effectiveness factors of the proposed national educational assessment model.

9.3.1 National Education Policy Perspectives on External Achievement Stimuli

The focus of this section was to collect national educational policy perspectives from policymakers at the Ministry of Education on the following external educational achievement stimuli: parental involvement; choice of schools; school privatisation; school autonomy; school accountability; and output financing.

An emerging theme in this section was that the policymakers were aware of the external educational achievement factors and were in favour of having the factors applied in Bhutanese schools. However, the policymakers identified challenges in capitalising on the factors because of various obstacles, albeit perceived, specific to Bhutan.

9.3.1.1 How important is it for parents to know about the performance of schools in which they plan to enrol their children?

Overall, the focus group perceived that the parental involvement in children's education was minimal, and that the parents viewed the children's education as the responsibility of schools. The group's view was evident in the following comment (Member B), "Parents feel that once their children are in schools, it is the responsibility of the schools to take care of their children". Further, the group (prompted by Member B) stated, "Parents should play a role of equal partner to teachers or anyone involved in their children's education". The focus group emphasized the need for policy interventions aimed at increasing parental involvement in children's education.

> There should be a model shift in the way parents are involved in our schools. Parents should be involved more in their children's physical, moral, and academic developments than their current way of being involving only in procuring children's stationery and participating in school management board meetings and other non-academic activities. (Member B)

The focus group (Member A) recognized the importance for parents to be aware of the performance of the schools their children were going to attend as evident in one statement, "Very important and there is no doubt about it". However, the group, revealed the challenges in making information about the performance of the schools accessible and comprehensible to the parents. As noted by one group member:

> We have not reached that situation where parents can assess how schools are performing. Some schools provide information through school calendars and school magazines, which is neither sufficient nor comprehensive for judging the school performance. (Member A)

Discussion

In line with the literature (Fan & Chen, 2001; Hoover-Dempsey, et al., 2001; Hoover-Dempsey & Sandler, 1995, 1997; Jeynes, 2007; Ritblatt, et al., 2002; Yan & Lin, 2005), the focus group highlighted the need for schools to connect with parents by sharing information about their children's performance and by involving them in school activities. Such a relationship between parents and schools has the potential for the latter to get greater support from the former in areas related to student learning.

9.3.1.2 What are the likely sources of information available to parents to learn about prospective schools for their children?

The focus group pointed out a range of sources as stated in the statement below:

> The likely sources of information are school magazines, school brochures, school calendars, school performance reports, and examination topper list. The Ministry is also in the process of finalizing management of information systems that will provide comparative information on school performance. Other sources of information are newspapers.
>
> (Focus Group)

Despite a range of potential sources of information available for the parents, the focus group was sceptical about the parents' ability to use them for evaluating the quality of their children's prospective schools. A statement from the focus group (Member A) indicates this, "All parents cannot read the information provided by schools because most of them are illiterate".

Discussion

With school magazines, brochures, calendars, student performance reports, and media referred to as popular sources of information for parents about their children's schools, the group perceived that the majority of parents may not be able to absorb this information because of Bhutan's low national literacy rate. The inability of most parents to use available print materials limits communication between parents and schools to oral communication. As reported in Ritblatt, et al. (2002), the lack of proper communication seems to be a key barrier to parental involvement in the activities of Bhutanese schools. The reliance on oral communication also raises doubts about the parents' ability to assist their children with homework and other academic activities, which is a challenge being faced by parents with a low SES and less formal education (Epstein, 1995; Epstein & Sanders, 2006). However, because the parental involvement in school activities or children's learning spans beyond academic activities in the forms of parenting, volunteering, decision making, and collaboration with communities (Epstein & Sanders, 2006), alternative school outreach programmes should be able to involve parents in school activities.

9.3.1.3 How would the publicity of students' achievement results, post-school enrolments in colleges, and receipt of scholarships affect school performance?

As expected, the focus group had positive views about the ways of recognizing students' academic achievement. The following comment highlights this view:

> "I think for class 12, scholarships make students work hard. In fact, a scholarship is one of the main incentives for students to work hard.... Scholarships foster competition among schools and students. This may make schools and students work hard. At times schools, especially private schools, tend to be judged based on the number of students who qualify for further education in colleges". (Member C)

However, the focus group was mindful of the adverse effects of an excessive emphasis placed on student achievement,"... too much focus on students' performance in examination often results in neglecting other aspects of wholesome education or holistic education".

The focus group also indicated a change in the future capacity of the Ministry of Education on the matters related to the question. A comment from the group indicates this:

> "Individual schools usually keep a record of the number of students who qualified for further studies in colleges. Such a record is not available at the national level. Education Management and Information System (EMIS), once completed, will provide information about students' post-school education and their previous schools". (Member B)

Discussion

The group perceived that the publicizing of student achievement results, post-school enrolments in colleges, and receipt of scholarships would make students work harder as they compete for scholarships. The group believed that hard work by students would result in improved student performance. However, the group was aware of the negative aspects of school accountability when it is driven by school league tables, as reported in Goldstein and Leckie (2008).

9.3.1.4 What are some of the potential benefits of allowing students and parents to select schools?

The group perceived parents' satisfaction and easing parents' socio-economic concerns as a couple of benefits from allowing students and parents to select schools:

> “The potential benefits will be making the parents satisfied with why they want to send their children to a particular school. Behind the satisfaction, there must be a variety of reasons, such as social and economic ones. Therefore, allowing students and parents to select schools will solve parents’ socio-economic problems”. (Member A)

This view is further supported by another member:

> In the past, many students from Thimphu used to seek admission in Zemgang High School, although the school was located in a socially backward part of the country. The reason for such interest in Zemgang High School was that the parents of those students hoped to solve some of the urban-related problems picked up by their children”. (Member C))

Parents’ desire to get their children admitted in schools closer to their residence also featured as a benefit, as implied by the comment (Member B), “Parents tend to opt for schools near to where they live. Therefore, parents often do not think about how schools are performing”.

Notwithstanding the perceived benefits from parents being able to choose schools for their children, the group (Member A) raised some negative or troublesome consequences of allowing parents to choose schools, “Giving choice to parents will result in concentrating students in high performing schools and fewer enrolments in low performing schools”.

In addition, the parents in Bhutan had little control over which school their children might attend as indicated by the group’s comment (Member B), “There is an assumption that parents have a choice of schools, but it is not like that. It is largely related to what District Education Officers plan and decide”. Further, the group commented, “...as it is the government who provides educational goods and services, parents have to comply with educational goods and services provided by the government, including which school their children should enrol”.

Discussion

As reported in Waslander, Pater and Weide (2010), parents’ satisfaction and reduced socio-economic concerns were the two benefits that the group felt would ensue from allowing parents to choose schools for their children. While the former is self-explanatory, the latter is related to students’ transition and school choice (Gibbons, et al., 2008; Schiller, 1999). However, the group mentioned that allowing parents to choose schools for their children would result in high-performing schools having more students than the low-performing schools, a perspective similar to the findings reported in the literature (Gibbons, et al., 2008; Gibbons & Telhaj, 2007; Hanushek, Kain, & Rivkin, 2004). Notwithstanding the merits and demerits of parents being able to choose schools for their children, the group stated that since schools were funded by the Royal Government of Bhutan, it is the prerogative of the government to allocate schools for students if it wishes, avoiding the possibility of school choice in Bhutan.

9.3.1.5 How important is it for schools to have freedom in procuring teaching and learning resources?

In principle, the focus group recognised the benefits of granting autonomy to schools in procuring teaching and learning materials as evident in the comment (Member B), "It is very important, especially for result-oriented management or outcome-based financing model. If schools were made totally responsible for holistic development of students, it is very important and critical that schools have autonomy in procuring teaching and learning resources".

However, what was considered potentially helpful for the schools in principle was not applied in the schools. A comment from the group confirms this (Member B), "...due to some constraints, such as inadequate staff for, and right expertise in, dealing with procurement matters, the Ministry of Education has decided not to grant such autonomy to schools".

Discussion

Granting autonomy to schools in procuring teaching and learning materials was perceived to benefit the schools, but the schools were not granted such autonomy because of the perception that schools were not ready. Given that the focus group felt that granting autonomy to schools in procuring teaching and learning materials would be beneficial, which is also in line with the literature (Fuchs & Wobmann, 2007; Maslowski, et al., 2007; OECD, 2010; West, et al., 2010; Wobmann, 2007), studies ought to be initiated to evaluate the feasibility of granting such autonomy to schools.

9.3.1.6 How important is it for schools to have freedom in selecting, appointing, and terminating teachers and staff?

The focus group recognised the importance of autonomy to schools for selecting, appointing, and terminating teachers and staff in its statement (Member A), "It is one of the highest degrees of autonomy. If it is doable, it is good and very important, but is it very practical now?" Alternatively, "These [freedoms indicating school autonomy] are possible in Bhutanese schools if all schools have minimum facilities already in place so that they have a level-playing field".

The focus group also raised some additional benefits from granting autonomy to schools. The group pointed out that school autonomy would make teachers feel closer and directly accountable to their schools, and that school autonomy would take care of the teacher deployment responsibility of the Ministry of Education.

> "If we look at it positively, one of the major headaches that the Ministry has is the deployment of teachers. Every year we get hundreds of requests for teachers. These teachers, as they work for the Ministry, feel more accountable to the Ministry than to their schools. If the schools select, deploy, and terminate teachers, they will feel accountable to schools". (Member B)

However, the group was quick to point out that school autonomy would lead to social inequalities by way of creating a rural-urban divide among students:

> "...in a situation where we have limited resources to offer teachers and the fact that the teacher salary is uniform across schools, it may be difficult to justify this kind of school autonomy... this kind of autonomy will broaden the rural-urban divide. With government policy interventions, such a rural-urban divide can be avoided and a uniform socio-economic growth can be achieved by allowing parents with equal access to uniform educational facilities across the country". (Member B)

Discussion

Empowering schools to recruit and terminate teachers and other school staff would make the teachers and the staff more accountable to schools and free the Ministry of Education from the responsibility of teacher recruitment and deployment. However, these benefits were obscured by the prospect of fostering inequalities between urban and rural schools because only a few good teachers would prefer to work in rural schools, resulting in differentiated learning opportunities in the two sectors. As suggested by Grauwe (2005), policymakers may grant autonomy to schools in conjunction with supportive strategies to develop the capacities of principals, teachers, and communities with a clear focus on autonomy as a tool for improving school performance and establishing equity in the school education system.

9.3.1.7 How important is it for schools to be able to use their budgets according to their needs and priorities?

The focus group indicated that it was up to schools to use their budgets in accordance with their plans:

> "It is very important. Schools should use their budgets where most needed and give the best results. Schools have to find out the factors related to the desired results and the budget should be used in developing these factors. It is up to the schools to take decisions on how they use their budgets". (Member A)

Discussion

The group underlined the importance of schools being able to use their budget where it is most needed, noting that it was up to the schools to decide how to use their budget. However, an examination of a school budget plan reveals that schools have little discretionary power in using their prescribed budget because of its strict bureaucratic clearance compliance and mandatory audit trailing requirements. Such elements in a budget have been known to obstruct schools in carrying out educational innovations to improve student achievement, especially improving the performance of the at-risk students (Timar & Roza, 2010).

9.3.1.8 If you had the authority to relate school performance to school funding, what would be your primary recommendation?

The possibility for financing Bhutanese schools based on their performance is unlikely for a number of reasons:

> "I think we should have the balance of head and heart. The heart may say this is good and go ahead, but is this practicable in the present situation? We need to be rational. I would not really have a lopsided recommendation that relating school performance to school financing is good". (Member A)

A reason for the performance-based financing being lopsided was the lack of comprehensive and reliable assessment practices that take into account students' socio-economic backgrounds, school resources, schools' locale, and students' holistic development:

> "Schools differ in terms of infrastructural facilities, teaching and learning resources, and communities from where students come. Most of the parents in urban schools are literate, while most of the parents in rural places are illiterate. Examinations might not actually tell you what the status of education or the quality of education is in schools, thereby [limiting the whole range of] the school performance". (Member A)

The group also feared the prospect of malpractices by schools to raise their performance to secure financial rewards (Member B), "We do not want teachers conducting examinations with answers written on the board. If school finance is linked to school performance and teachers conduct examinations, it is obvious that such negative behaviours are bound to happen".

However, output-based financing, is largely viewed as more efficient than input-based financing; and consequently, the focus group outlined their preference for the former.

> "We want to finance schools based on their achievement targets, rather than basing the finance on the resources related to inputs only. Targets would be based on students' all round development and their post-school achievements. We are planning to develop outcome-based financing in the near future". (Member B)

Discussion

The group did not favour the idea of performance-based funding because the Bhutanese education system did not have the capability to develop a comprehensive list of school performance indicators other than the examination-oriented ones. This statement is difficult to interpret, but it reveals a gap between the school fiscal policies and school performances. In addition, the statement seems to overlook some of the well-established aspects of performance-based funding in the Bhutanese education system such as career ladder, meritorious promotion, master teacher, and in-service training (Foster & Marquart, 1984). The gap, if it existed, would mean that the preceding aspects of performance-based funding are followed as routines (e.g., promoting teachers based on some kind of periodic schedule) in Bhutanese schools rather than as tools for improving student outcomes (Foster & Marquart, 1984). Because there was a desire for performance-based funding of the schools, the Ministry of Education may care to explore suitable models for instituting performance-based funding.

9.3.2 National Education Policy on Enabling Learning Environments

This section explores the national education policy perspectives on enabling learning environments in schools. An emerging theme in the section is that the attributes of enabling learning environments are viewed largely as being related to physical infrastructures (e.g., electricity, water supply) of schools. A few intangible attributes (e.g., school leadership) also figured on the focus group's list of the attributes of enabling learning environments in schools.

9.3.2.1 Thinking about some attributes of an effective learning environment for schools, how many such factors can you promptly list?

The focus group raised the following as the attributes of effective learning environments: teaching and learning materials; career counselling; electricity; road access; the Internet; water supply; library books; teacher-pupil ratio; classroom size; school leadership; and health facilities.

9.3.2.2 How do the elements (listed in the previous question) contribute to building an education friendly atmosphere in schools?

The group stated that the elements would foster a child-friendly school environment and child-centred teaching and learning activities in schools (Member B). "We would want the classroom climate to be conducive, encouraging enquiry, not rote learning. Children should be encouraged to stand up and ask questions when they have doubts, rather than holding back. Education should be child-centred".

9.3.2.3 How would you relate the elements of an education-friendly atmosphere to student performance?

The group perceived a positive relationship between the elements of an education friendly atmosphere and student performance.

> "If you have got a classroom where we have good student and teacher interactions, minimum class size, motivated teachers with good content knowledge and pedagogical skills, good leadership, a school with such classrooms will be a very happy place to be for children. The children in such a school will learn with enjoyment which will have a positive and direct impact on their learning, and hence their school performance." (Member B)

The group also raised the importance of career counselling services in schools (Member B), "Career counselling or school-life linkages play a very important role. If a child is educated at the right age on what their aptitudes, goals, interests, and efforts are, school can become more meaningful to the child's future employment".

Discussion

As expected, physical infrastructural resources are viewed as important for developing enabling learning environments in schools. However, equally important, if not more, are some intangible constructs such as school climate and student engagement. The focus group overlooked the intangible constructs in its list of the attributes of enabling learning environments in schools. This finding indicates that the Ministry of Education might not have a comprehensive policy on the attributes of enabling learning environments in schools, but should develop such a policy.

The focus group's perspective of the relation between an enabling learning environment and student achievement suggests that adequate physical infrastructural resources will develop enabling learning environments in schools and eventually result in child-centred education. This finding suggests that physical infrastructural resources and other intangible attributes of enabling learning environments are compensatory; therefore, focussing on the former compensates for the latter. However, the literature shows that physical infrastructural resources and other intangible attributes of enabling learning environments are not always compensatory, but are rather disjunctive, indicating the need for national educational policies to have equal emphasis on physical infrastructural resources and other intangible attributes of enabling learning environments (Barber & Mourshed, 2007; Hanushek, 1997; Hoy & Hannum, 1997; OECD, 2004a; Wilson, 2004). This factor should be reflected in the development of policy suggested in the previous paragraph.

9.3.3 Mechanisms for Evaluating National Education Policies

This section explores the views of the focus group about the practices of evaluating national education policies. An emerging theme in this section is that the standard practices of policy formulation seemed to have been followed in developing national education policies, but the policies were not then evaluated.

9.3.3.1 Based on your experience, what are the steps you usually follow in formulating policies on educational programmes?

Most of the standard steps involved in formulating policies were inherent in the group members' experience with policy formulations:

> "...identification of issues or problems is very important. ...next is getting budgets for solving the problem. ...getting to solve the problem and involving stakeholders from various agencies are other steps. ...we invite people from various departments to identify issues or problems related to their goals and needs and relate their problems to policy guidelines and then frame programmes for them". (Member A)

9.3.3.2 How many times were you involved in evaluating national education policies?

Few group members had experience with evaluating educational policies (Member B), "I do not remember being involved in evaluating education policies". On the other hand, the group contested the validity of the evaluation if policy developers were involved in evaluating their own policies (Member B), "...being policy developers, how relevant it is for us to evaluate our own policies is another issue".

Discussion

The experience and the mentioning of the sequential steps of standard mechanisms of policy formulation (Bridgman & Davis, 2007) is indicative of the mechanisms being followed in developing national education policies in Bhutan. However, the lack of experience of the focus group in evaluating national education policies indicates that the policies, once developed and implemented, are not evaluated. As stated by Bridgman and Davis (2007), not evaluating the policies would result in the lack of knowledge about: how the policies have achieved their objectives; who to hold accountable for the implementation of the policies; and the directions and clues for future policy making. These implications are serious because they suggest the lack of policy analyses and the loss of opportunities for policy learning, which need to be addressed in formulating national education policies and their later development(Parsons, 1995).

9.3.4 Miscellaneous: Further Perspectives

This section presents the focus group's perspectives that were not referred to in the earlier questions, but elicited through the question, "Is there anything further you would like to say on external stimuli for student achievement?" Three points were mentioned by the group: a need for enabling conditions for applying the law of supply and demand between the colleges of teacher education and the Ministry of Education; a need for the monitoring agency to be autonomous from the organizations it monitors; and the relationship between student graduates and labour markets.

First, the group pointed out the need for enabling conditions to apply the law of supply and demand to deliver accountability to the colleges of teacher education:

> "The role of the Ministry of Education being responsible for the colleges of education and responsible for recruiting and deploying graduate teachers makes the law of supply and demand difficult to apply. The Ministry of Education has to take in all graduate teachers from the colleges of education". (Member B)

The focus group indicated a change in the way the Ministry of Education will recruit teacher graduates from the colleges of teacher education:

> "Now we are trying to say that we change the role of the Ministry of Education. We are the client and we will demand the type of services we want, the quality of teachers we want, and the quantity of teachers we want from the colleges of education. If we do not get what we want, then we can move to another vendor, in Bhutan or outside Bhutan. So we need to change that role and it is very important". (Member B)

Second, the group raised the need for the monitoring organization to be autonomous from the organizations it monitors.

> "If we look at the Department of School Education, it is the primary department for transaction of educational activities. Its activities are monitored by Education Monitoring and Support Service Division of the Department of Education. In the functioning of good governance, we cannot have doing and monitoring by the same agency.... Now, we are saying that everybody should play the right role, so the monitoring agency should be independent from the organizations it monitors". (Member B)

Third, the group perceived a mismatch between the labour market and the school curriculum, as underlined by the comment (Member B), “Students coming out of schools and institutions find themselves very little equipped for the labour market”.

Discussion

These three points indicate the challenges that the Ministry of Education has been facing. The first point indicates that the teacher graduates from the colleges of teacher education were guaranteed employment by the Ministry, and this guarantee was viewed as compromising the quality of teachers. In addition, the practice was also seen as protecting the colleges of teacher education from the consequences related to any lack of preparedness of their students for the teaching profession. As a result, the Ministry of Education seems to use market mechanisms to exert pressure on the colleges of education to produce quality teachers.

The second point indicates the challenges being faced by the monitoring agency in delivering accountability because of its responsibility to the agents it monitors. This is a serious issue because it risks biased results from the monitoring activities. There is a possibility for the monitoring agency to be interfered with by the agents it monitors. However, it seems that the Ministry of Education is aware of the issue and that proper steps are being taken to address the issue.

The third point indicates a gap between the school curriculum and the knowledge and skills required in the labour market. A further study should be initiated into the issue: perhaps a tracer study similar to that of Hanushek and Wobmann (2006) might be conducted.

9.4 **Summary and Implications for Policy**

A broad range of policy implications can be drawn from the focus group interview findings.

The group recognised the importance of disseminating adequate and proper information to parents about prospective schools for their children. However, the value of such information was perceived to be contentious because of Bhutan’s low literacy rate. A policy implication from this might be to explore alternative ways of disseminating information. This might include the Ministry’s Educational Monitoring and Information System and the use of the National Language as the medium, coupled with the range of conventional information sources for parents to know about prospective schools for their children.

Focus group data revealed the significance of publishing student achievement results, post-student enrolments in colleges, and receipt of student scholarships. The analyses also revealed that programme officers were aware of possible adverse effects of student achievement-oriented publicity. This finding suggests that relevant policies might well be formulated on student achievement-oriented publicity that encourages student learning and discourages adaptive malpractices in schools.

The group foresaw some benefits from allowing parents to choose schools for their children. Parents' satisfaction and easing parents' socio-economic concerns emerged as two major advantages that may arise from allowing parents to select schools for their children. On the other hand, the availability of options for parents to select schools for their children was contested given its potential to heighten the present rural-urban enrolment imbalance, with the parents preferring to admit their children to urban schools where teachers are better resourced than those in rural schools. However, parents have little freedom to exercise preferences for schools for their children because the Royal Government of Bhutan allocates places. These findings suggest that the benefits of allowing parents to choose schools for their children be viewed from the accountability aspects and that mechanisms to prevent a rural-urban enrolment divide be explored through relevant studies.

The group recognised the importance of granting autonomy to schools for procuring teaching and learning materials. Contrary to its perceived importance, the schools still lacked this autonomy. A policy implication from the finding is that studies ought to be initiated to explore a range of workable options for granting autonomy to schools for procuring teaching and learning resources. The group was aware of the benefits of granting autonomy to the schools for recruiting and terminating teachers and other staff. However, schools were not granted such autonomy. This finding suggests that studies be carried out to investigate strategies or processes whereby schools might realistically recruit teachers, with the aim of providing greater responsibility to schools in developing and managing school human resources.

Currently Bhutanese principals have complete autonomy in the disposal of their allocated budget. A policy implication from the finding is that policy guidelines be developed for training school personnel in budgeting and financial management.

The future implementation of output-based financing in Bhutanese schools is unlikely, as the group currently viewed it as a dubious challenge rather than as a means of improving school accountability, suggesting the need for a feasibility study on the prospect of introducing output-based financing in Bhutanese schools. In other words, there is a need to investigate the negative and positive effects of output-based financing and prevailing attitudes towards it.

The group was aware of the complementary relationship between an effective learning environment and student learning. Further, the group endorsed the importance of ensuring that the elements of positive learning environments existed in schools. This finding suggests that proper monitoring and assessment tools be developed to monitor and evaluate the health of the school environment.

Individuals in the group had considerable experience in formulating policies. However, the group had little experience in evaluating them. An implication from this finding is that a mechanism be developed that extends policy formulation to policy evaluation, and include consideration of appropriate structures and processes for this.

The last section showed that the group was not happy with the lack of enabling conditions for supply and demand to regulate colleges of teacher education, and the lack of autonomy of the monitoring agent from the circumstances monitored. Some implications from this finding are that studies be conducted into teacher recruitment and the relationship between monitoring and monitored organisations.

9.5 Chapter Conclusion

This chapter presented the perspectives of the key policy makers in the Bhutanese education system on external achievement stimuli, effective learning environments, and evaluation of national education policies. With respect to external achievement stimuli, responses to the questions revealed that the policy makers endorsed the importance of recognising student achievement through publicity and other meritorious awards. On the other hand, the responses to the questions revealed that the policies needed to be formulated on making schools more accountable to the stakeholders through granting them greater autonomy. Responses to the questions on effective learning environments revealed that the policy makers recognise the relationship between effective learning environments and student learning, though most of the elements discussed were related to the material aspect of school resources. Finally, responses to the questions on evaluation of education policies revealed that the policymakers had considerable experience in developing education policies, but had little experience in evaluating education policies.

Chapter 10

SUMMARY AND DISCUSSION

10.1 Introduction

This chapter restates the aim of the study and summarises the literature review and the methods used. The chapter then discusses the main themes emerging from the analyses of data collected from the different levels of the proposed national educational assessment model, followed by a discussion of the interplay of the findings across the levels.

10.2 Summary of the Aim of the Study (Chapter 1)

The aim of this study has been to develop a national educational assessment model for Bhutan. As reported in Section 1.3 of Chapter 1, ten outcomes have been derived from this aim, covering a range of student, teacher, and school characteristics.

10.3 Summary of the Literature Review and the Research Methods (Chapters 2, 3, and 4)

A review of a range of studies in Chapters 2 and 3 guided the development of a national educational assessment model for the Bhutanese education system. Following the development of the model, different methods of collecting data for the model were reviewed in Chapter 4.

A review, in Chapter 2, of the literature on current international educational assessments such as NAEP, TIMSS, and PISA resulted in the identification of designs, principles, and purposes for developing a national educational assessment model for the Bhutanese education system. Some of the design elements to develop the proposed national educational assessment model included: consensual approaches; low-stakes, high-value status; cognitive and non-cognitive aspects of teaching and learning; and descriptive achievement levels. The bedrock of the purposes of the proposed model was to provide stakeholders in the Bhutanese education system with research-based knowledge about diverse educational effectiveness factors to help them improve the quality of school education. Like NAEP, TIMSS, and PISA, the proposed model was a cross-sectional survey, with the participants sampled by a two-stage stratified cluster sampling method. The review also revealed diverse critiques on the validity of NAEP, TIMSS, and NAEP as assessments of, or for, student learning. The misalignment of NAEP, TIMSS, and NAEP with the curriculum followed by their participant schools often served as a point of departure for the critiques (Dohn, 2007; Jenkins, 2000; Nardi, 2008; Orpwood, 2000; Prais, 2003). However, the review also provided methods to ensure that large-scale assessments like NAEP, TIMSS, and PISA are aligned with the curriculum of participant schools (Tshering & Prain, 2011; Webb, 1997, 1999, 2006).

The next phase of the review, in Chapter 3, guided the building of a national educational assessment model for the Bhutanese education system. Different traditions of educational effectiveness studies informed the development of a multi-level skeletal national educational assessment model for the Bhutanese education system. One of the salient properties of this model is its capacity to include students, teachers, schools, and systemic context as change agents of student achievement. The skeletal model is then fleshed out with student, teacher, school, and systemic context characteristics that are known to correlate with student achievement. PISA, TIMSS, and NAEP provided a pool of such characteristics, and guided the subsequent in-depth review of the characteristics as factors of student achievement. The complete national educational assessment model for the Bhutanese education system is presented in Section 3.5.1 of Chapter 3. The chapter also described the procedures followed in the lead-up to the endorsement of the model by stakeholders in the Bhutanese education system.

Chapter 4 described the research methods used in this study. The core elements of the research methods of this study sit within a pragmatic paradigm, using mixed methods, cross-sectional survey, focus group interview, stratified two-stage cluster sampling design, and purposeful sampling design. The chapter also described the procedures followed in developing and in-field administration of the Mathematics test, the student questionnaire, the teacher questionnaire, the school questionnaire, and the focus group interview. A range of simple and scale indices, together with their validation procedures, was also described. The indices were made ready for further analyses in Chapters 5 through 9. The results from the analyses revealed emerging themes with potential to guide educational policy decisions and interventions for improving the Bhutanese education system. These themes will now be discussed on a chapter-by-chapter basis.

10.4 Emerging Themes from Analyses of the Mathematics Test (Chapter 5)

It was assumed that the following knowledge about student achievement in the Mathematics test would provide a basis for improving the quality of the Bhutanese education system: the profiles of students’ mathematical knowledge and skills; the profiles of students’ higher-order thinking skills; the students’ achievement levels at school, district, and region; the equity in, and accessibility to, educational resources; and the international benchmarks.

This study showed that profiles of students’ mathematical knowledge and skills in the six strands of the Bhutanese Grade 10 mathematics curriculum clustered around Level 1 and Level 2 of the PISA Mathematics Proficiency Scale, with one quarter of the students performing below Level 1. However, the cumulative mathematical knowledge and skills expected of Grade 10 students, defined as learning outcomes, by the Bhutanese Grade 10 mathematics curriculum (CAPSD & BBE, 2007) mostly corresponded to Level 5 and Level 6 of the PISA Mathematics Proficiency Scale (OECD, 2004a). This adverse difference between the expected learning outcomes and the actual attainment levels of the Grade 10 students suggests the lack of success of the curriculum goals and objectives by teachers and students. The profiles of students' thinking skills also indicate that the students need to develop higher-order thinking skills, suggesting limited opportunities for the students to acquire and develop higher-order thinking skills in the classroom. The non-attainment of the curriculum goals and objectives may also be one of the factors that contributed to Bhutan's benchmarks being similar to the international benchmarks of countries that performed relatively poorly in PISA (OECD, 2004a).

As mentioned in Chapter 3, student achievement is determined by a range of factors related to students, teachers, schools, and the education system. Therefore, the following sections discuss many such factors affecting the performance of Grade 10 Bhutanese students on the Mathematics test, as revealed by this study.

10.5 Emerging Themes from Analyses of the Student Questionnaire (Chapter 6)

Educational effectiveness factors at student level comprised the following: gender; age; socio-economic status (SES); engagement; motivation; self-beliefs; self-regulated learning skills; homework; ICT; classroom management; and school climate. Overall, this study showed that these factors are associated with student achievement.

Students' gender, as a factor of achievement, has been widely reported in the literature (Clark, Lee, et al., 2008; Clark, Thompson, et al., 2008; Lisle, et al., 2005; Younger & Warrington, 2007). Because Bhutan provides equal learning opportunities to students through a co-education system, gendered achievement difference is less likely. However, this study showed that on average boys outperformed girls on the Mathematics test, suggesting the role of students' gender in achievement. While gendered achievement difference depends on school subjects, the finding shows the need for Bhutanese schools to develop educational interventions to engage girls in learning mathematics. The literature provides diverse areas of educational interventions for reducing gendered achievement difference: peer influence (Crosnoe, et al., 2008); co-education (Malacova, 2007; Younger & Warrington, 2006); gendered achievement orientations (Houtte, 2004a, 2004b; J. L. Smith, 2006); student motivation and attitudes (Gaer, et al., 2006); students' choice of school subjects (Cox, 2005). Some of these areas may help the Bhutanese education system to reduce the gendered achievement difference in mathematics.

Analyses of data on students' age showed that younger students performed better than older students, and that the average age of the students at Grade 10 was greater than the expected average age. While the superior performance of the relatively younger students suggests congruity between students' age and curriculum goals, the high average age suggests grade retention. On average, a quarter of the students reported repeating one or more grades before attaining Grade 10. It is plausible that the students were retained in the same grade for an additional year to make them attain the required level of curriculum goals and objectives. However, such practices have implications. First, research has consistently shown that retaining students in the same grade does not result in achievement gains (Braymen & Piersel, 1987; Cameron & Wilson, 1990; Jimerson, 2001; Jimerson, et al., 2002; Jimerson, et al., 2006). Second, grade retention results in the loss of scarce educational resources that could have been invested in other students. The retained students are also disadvantaged at employment opportunities compared with other students, with the opportunity costs of the retained students rising as they complete their education (Eide & Goldhaber, 2005; OECD, 2004a). Third, the practice of grade retention also implies that the teachers view students as solely responsible for learning. The literature suggests that grade retention be substituted by appropriate instructional strategies, improved curriculum design, and remedial lessons to improve student learning (Braymen & Piersel, 1987; Cameron & Wilson, 1990; Hauck & Finch(Jr), 1993; Jimerson, 2001; Kundert, et al., 1995).

Students' SES played a significant role in their performance on the Mathematics test. As commonly reported in the literature, on average, high-SES students outperformed low-SES students (Caldas & Bankston(III), 1997; Lee, et al., 2007; Mullis, et al., 2004; OECD, 2004a, 2007a; Sirin, 2005; White, 1982). This finding suggests that students' SES be used as one of the criteria for allocating educational resources to schools. The mean of students' SES is usually reported as the schools' SES (OECD, 2010). With their SES known, schools can be ranked in terms of their SES, and resources can be distributed inversely to their ranking so that schools with low SES get more resources.

Students' motivation, self-beliefs, self-regulated learning strategies, and learning preferences revealed a consistent pattern in the teaching and learning processes taking place in Bhutanese schools. Instrumental motivation was greater than intrinsic motivation. The literature relates instrumental motivation to performance goal-orientation which is characterised as short-lived and impressionistic (Ames, 1992; Ames & Archer, 1988; Bell & Kozlowski, 2002; Dweck, 1986; Hidi & Harackiewickz, 2000), implying that the students viewed teaching and learning processes as confirming teachers' preferences and requirements rather than viewing them as ways to develop knowledge and skills necessary for life-long learning.

Students' self-beliefs showed that most of the students had high self-efficacy about mathematics. However, approximately one-half of the students had a low self-concept and a high anxiety about mathematics. Because self-concept has been linked with students' experiences accrued from social comparisons and self-descriptions (Bong & Clark, 1999), a low-self-concept suggests the lack of relevant opportunities for students to develop a positive perspective. Similarly, subject anxiety has been linked to teaching strategies (Alsup, 2004; Furner & Berman, 2003; Hellum-Alexander, 2010; Norwood, 1994), and classroom climate (Ma, 1999), indicating that most of the students did not experience appropriate teaching strategies and classroom climate.

The analyses of data on the students' self-regulatory learning strategies showed that the students used diverse self-regulatory learning strategies. From the three dimensions of self-regulatory strategies, only the control strategy resulted in significant improvement in student achievement. Although statistically not significant, the memorisation and the elaboration strategies seemed to correlate negatively with student achievement in mathematics. Despite its statistically non-significant association with student achievement, memorisation appeared to be a popular strategy used by Bhutanese students to learn mathematics. This has implications for students' future careers or further studies. Self-regulated learning strategies such as self-monitoring and elaboration strategies have been cited as characteristics of students with the potential for life-long learning (Cropley, 1981; Knapper & Cropley, 2000) and the potential for full participation in increasingly complex labour markets (OECD, 2004d). On the other hand, students' use of a memorisation strategy implies the prevalence of traditional teaching and learning approaches in schools (Boekaerts & Niemivirta, 2005).

Results from the analyses of data on students' preferences for learning environments showed that the students preferred competitive learning environments more than cooperative learning environments. In addition, students who reported a higher preference for competitive learning environments performed better on the Mathematics test than students who reported a lower preference for the same. These findings have implications on students' goal orientations and teachers' classroom practices. Competitive learning preferences suggest that students were extrinsically motivated and driven by performance goals (Covington & Omelich, 1984) and that teachers used traditional teaching strategies in the classroom (Covington & Omelich, 1984; Ediger, 1996). As stated by Covington and Omelich (1984) and Stapel and Koomen (2005), students' strong preference for competitive learning environments, which encourage students to count their success on the challenges of their colleagues, may also be attributed to norm-based assessments that is widely practised in the Bhutanese education system.

Homework has been perceived as a correlate of student achievement (Brock, et al., 2007; Cooper, et al., 2006; Epstein & VanVoorhis, 2001; Muhlenbruck, et al., 2000). Corroborating this view, this study showed that homework correlated with student achievement. However, the relation between homework and student achievement was moderated by other factors. Similar to the findings reported in the literature (Cooper & Valentine, 2001), the relation peaked at a certain frequency and duration of homework. This finding suggests that schools need to consider frequency and duration of homework when developing homework policies.

The role of ICT in mainstream schools is to improve student learning by assisting them to access information about their classroom lessons or homework (Lim, 2002; Ruthven, et al., 2004). Students' experience with ICT, places where students access ICT, the duration of students' use of ICT, students' proficiency in performing ICT functions, students' confidence in ICT, and students' attitude towards ICT have been known to determine the impact of ICT on student learning (Martin, Mullis, Gonzalez, & Chrostowski, 2004; OECD, 2004b). This study showed that more than one-half of Grade 10 Bhutanese students have never used ICT. Of the students who used ICT, approximately one-quarter of the students reported using ICT at places other than schools. These findings indicate the need for the Ministry of Education to increase its investment in ICT infrastructure in schools. The duration for which students have been using ICT also suggests that ICT is a recent educational innovation in Bhutanese schools because the majority of students reported using ICT for three or fewer years. Despite ICT being recent, students reported using diverse ICT functions relating to the Internet and software applications. However, fewer than one-half of students reported using any one of the ICT functions. These findings suggest that the classroom lessons were not ICT-oriented or ICT-enabling, resulting in the poor use of ICT as a mediational tool by students (Drent & Meelissen, 2008). In addition, students were more confident in performing routine tasks compared to either Internet tasks or high-level tasks, suggesting less use of the Internet or advanced software by students for furthering their understanding of classroom lessons. Notwithstanding their low ICT confidence, the majority of students had a positive attitude towards ICT and its functions.

Effective classroom management sets the stage for a safe and functional class, and determines the success of all teaching and learning activities that occur in the class. Consequently, effective classrooms should relate positively with student achievement. The majority of students reported much teacher support and few disciplinary problems in the class, indicating the prevalence of effective classroom management (Cothran, et al., 2003; Marzano, 2003a; Simonsen, et al., 2008). However, teacher support as a dimension of effective classroom management did not result in significant change in student achievement, contradicting the findings reported in Marzano (2003a).

One possible reason for the contradictory finding is that the students might have chosen to misidentify themselves with positive statements about their mathematics teachers.

An enabling school climate is expected to result in improved student achievement. The majority of students reported a positive attitude towards their schools, a strong sense of connectedness to their schools, and a favourable relationship with their teachers. As reported in the literature (Brand, et al., 2003; Marzano, 2001; Scheerens, 1990, 1997, 2000; Scheerens, et al., 2003), overall, this finding suggests the existence of an enabling school climate in Bhutanese schools. However, the minority of students who had less positive attitudes, less sense of connectedness, and fewer favourable relationships with teachers is indicative of the need for Bhutanese schools to nurture an enabling school climate.

10.6 Emerging Themes from Analyses of the Teacher Questionnaire (Chapter 7)

The following teacher-related educational effectiveness factors were expected to correlate with student achievement: demographic profile; professional collaboration; professional development; appraisal and feedback; school climate; beliefs about teaching approaches; use of teaching approaches; teaching constraints; self-efficacy; homework; tests and examinations; and calculators and computers. The analyses of teacher questionnaire data showed that these teacher-related educational effectiveness factors were related to student achievement, and that the teachers' differed in their perceptions of these factors.

Knowledge about the teachers' demographic profiles has the potential to guide stakeholders in the Bhutanese education system in developing educational interventions. Unlike most countries that participated in TIMSS (Mullis, et al., 2008), fewer than one-fourth of the mathematics teachers were female in the secondary schools. Given that a gendered teaching workforce has been related to gendered student achievement through students' role modelling, subjective evaluation, stereotyping, and classroom interactions (Dee, 2007; Ehrenberg, et al., 1995; Steele, 1997), the finding indicates the need for the Ministry of Education to devise policies aimed at achieving gender parity in the mathematics teacher workforce. Compared with the teaching workforce in other countries (Mullis, et al., 2008), the mathematics teacher workforce in Bhutan is relatively young. As reported in the OECD (2009a), having a young teaching workforce in the system may benefit the system in terms of salary costs, but having a matured teaching workforce in the system may also benefit the system in terms of its cumulative teaching experience (Mullis, et al., 2008). Given that the mathematics teacher workforce in the Bhutanese education system constitutes a range of age cohorts, the Ministry of Education could use teachers' age as one of the elements in its teacher deployment policies. In addition to their gender and age, teachers' educational attainment also indicated a need for the

Ministry of Education to develop strategic professional development programmes for its teachers. Two popular teacher qualifications reported by the mathematics teachers are B.Ed and PGCE. In line with the complementary relationship between teachers' subject-matter expertise and pedagogical competence (OECD, 2001), teachers with PGCE may need professional development in pedagogical knowledge and skills, while teachers with B.Ed may need professional development in subject knowledge and skills.

Professional activities of teachers in schools have been known to relate to higher teacher cooperation and higher student achievement (OECD, 2009a). Similar to the pattern reported in the OECD (2009a), this study showed that Bhutanese mathematics teachers collaborated more frequently in discussion of teaching strategies, discussion of mathematical concepts, and preparation of teaching materials than they did collaborating by observing one another's classes. While all forms of teacher collaborations contribute towards building enabling environments in schools, teacher collaboration in reciprocal observations of class has been viewed as a more progressive form of teacher collaboration because it is directly related to classroom practices (Clemant & Vandenberghe, 2000; OECD, 2009a). Therefore, schools may encourage reciprocal observations of classes by teachers. As reported in the OECD (2009a), schools should encourage this by allocating time for teachers to engage in reciprocal observation of one another's class with the objective of generating positive feedback and efficient teaching and learning strategies.

Professional development programmes for teachers are as important as the dynamic aspect of school curriculum. Teachers need to continuously update their content and pedagogical knowledge and skills to meet the challenges associated with a dynamic school curriculum, the diverse learning needs of students, and the increasing needs to reach out to parents (OECD, 2009a). This study showed that, on average, about 25% of Bhutanese Grade 10 mathematics teachers reported having participated in some form of professional development in the past two years. This rate of participation in professional development is considerably lower than the rate of teachers' participation in professional development reported in the OECD (2009a), which was as high as 89%. Despite the low rate of Bhutanese teachers' participation in professional development, on average, fewer than 20% of the teachers reported not needing any form of professional development. This suggests that the teachers faced difficulties in participating in professional development despite being aware of their need for professional development. Over one-fourth of the teachers reported a conflict between professional development schedules with their work schedule, and over one-half of the teachers reported a lack of suitable professional development as some reasons for not being able to participate in professional development. These findings indicate the need for proper planning related to teacher professional development in the Bhutanese education system. The cost of failing to provide the teachers with relevant and sustained professional development for Bhutan is implicit in the fact that the students of the teachers with less need for professional development performed better in the Mathematics test than the students of the teachers with greater need for professional development. This finding also supports the findings reported in Yoon, Duncan, Lee, Scarloss, and Shapley (2007).

Teacher appraisal, as a means for encouraging professional learning and growth in teachers, identifying opportunities for additional support for teachers, and providing a measure of accountability with the view to fostering sustained teacher development, has the potential to improve teaching and learning in schools. The findings from this study support this assumption. The students of teachers who reported that their principals viewed the different areas of appraisal with high importance outperformed the students of teachers who reported that their principals viewed the different areas of appraisal with low importance. Similar claims have been reported in the OECD (2009a). While it is encouraging to note that schools in Bhutan reported practising some forms of teacher appraisals, the difference in the emphasis made on the appraisal areas raises concerns. Because the appraisal areas are disjunctive, high impact on teaching and learning from greater emphasis on some areas of the appraisal cannot compensate for the low impact from lesser emphasis on some areas of the appraisal; each area of the appraisal needs equally high emphasis to bring about improvement in teaching and learning (OECD, 2009a). As reported in the literature (Casey, et al., 1997; Gratton, 2004; R. Smith, 1995), teacher appraisal in Bhutanese schools may be improved by improving schools' commitment, providing relevant support, developing and communicating appraisal areas, nurturing a positive culture of follow-up and feedback practices, and aligning the appraisal areas with school goals.

Effective classroom management is indispensible for a safe and functional class, and for all teaching and learning activities in a class to be successful (Cothran, et al., 2003). An orderly and respectful classroom climate is indicative of effective classroom management or fewer classroom disciplinary problems. The findings from this study showed that Bhutanese schools have fewer classroom disciplinary problems compared to schools in other countries. For example, the OECD (2009a) reported fewer than three-quarters of the teachers reporting that their students took care to create a pleasant learning atmosphere in the class, whereas this study showed almost nine in ten Bhutanese teachers reporting the same. Fewer classroom disciplinary problems in Bhutanese schools may be related to Bhutan's culture of respecting teachers as second parents. In addition, as the OECD (2009a) related fewer classroom disciplinary problems with constructivist views of teaching, it is likely that Bhutanese teachers believed in constructivist views of teaching. However, about one-quarter of Bhutanese teachers reported having faced some forms of classroom disciplinary problems. Because classroom disciplinary problems are related to student engagement as well as student achievement (Marzano, 2003a), educational interventions to improve classroom climate need to be re-enforced in Bhutanese schools. The literature (Doyle, 1980; Marzano, 2003a, 2003b; Simonsen, et al., 2008) has a range of research-based strategies, explicitly mentioned in Chapter 3, related to improving classroom climate which may be introduced in Bhutanese schools.

Knowledge about school climate reveals a valuable insight into the overall orderliness or ‘temperature’ of schools so that appropriate interventions can be developed to improve the areas related to school climate. Because school climate has been known to relate to student outcomes (C. S. Anderson, 1982; Marzano, 2001; Sammons, et al., 1995; Scheerens, 2000), improvement in school climate may result in improvement in student achievement. Overall, this study showed that the majority of Bhutanese Grade 10 students were taught by teachers who reported a positive school climate. This is encouraging because it suggests that most of the schools have enabling learning environments as characterised by high teacher satisfaction, high teacher competence, high teacher expectation for students, high parental involvement in school activities, and high student goals. However, the lack of a distinct pattern in the relation between student achievement and school climate in the Bhutanese context suggests that Bhutanese teachers might have either overstated their perspectives or other factors diffused the positive impacts of a good school climate. As controversial as the finding may be, the knowledge that Bhutanese teachers self-reported positively about school climate indicates the awareness of, and the readiness for, school climate-related educational interventions among the teachers. Therefore, opportunities could be provided for the teachers to participate in professional development in school climate.

Teachers’ beliefs about teaching predispose their teaching practices. While there is no single teaching strategy capable of bringing about an intended improvement in student learning (Harris, 1998; Marzano, et al., 2000), teaching approaches are commonly aligned with either direct transmission or constructivist views of teaching, with the latter being more preferred to the former (OECD, 2009a). Supporting the general trend, Bhutanese teachers also reported relatively strong beliefs in constructivist teaching approaches. However, student achievement did not relate to constructivist teaching approaches as strongly as it did to direct transmission teaching approaches. This indicates that the teachers might have endorsed constructivist views of teaching because it is perceived to be professionally desirable. Furthermore, although comparatively fewer, quite a number of teachers also endorsed direct transmission views of teaching. This finding shows that the teachers might be oscillating in their beliefs between constructivist and direct transmission views of teaching. As suggested in the OECD (2009a), it may be useful to develop the awareness of the difference between these views of teaching in the teachers during pre-service training and through professional development.

As discussed earlier, teachers' beliefs about teaching should determine their use of teaching strategies in the class. Despite a relatively strong endorsement of constructivist views of teaching by the teachers, the use of structured teaching strategies by Bhutanese teachers was more frequent than the use of student-oriented and enhanced teaching strategies. While this aligns with the findings reported in the OECD (2009a), the long term benefits of student-oriented and enhanced teaching strategies to students need to be highlighted. As life-long learning is the model of the Bhutanese education system, students should be able to organise and accomplish their studies with minimal support from their teachers so that they are able to continue their education independent of formal support from teachers. Such ability can be developed by teachers in the class using student-oriented and enhanced teaching strategies, which are known to promote active engagement of students in their learning, by teachers in the class (Marzano, 2001; OECD, 2009a). Therefore, greater emphasis may be made on orienting teachers towards practising student-oriented and enhanced teaching activities more frequently than structured teaching approaches. Teachers' ability to use a range of teaching approaches also depends on other factors such as student characteristics and material resources. This study showed that approximately one in four teachers reported having been constrained a lot from teaching effectively by students with different academic abilities, students with special needs, and uninterested students. In addition, about an equivalent number of teachers reported having been constrained from teaching effectively by the shortage of instructional materials. These findings suggest that not only the teachers need professional development, but the Ministry of Education also needs to study the best practices of allocating teaching and learning resources being followed in some of the best performing schools of the PISA countries.

Teachers' self-efficacy has been known to impact on their job satisfaction and student achievement (Caprara, et al., 2003; Goddard & Goddard, 2001; Goddard, et al., 2000; J. A. Ross, 1992; Tschannen-Moran, et al., 1998). Supporting this view, the findings from this study showed that the students taught by the teachers with high self-efficacy performed better than the students taught by the teachers with low self-efficacy. Furthermore, the majority of Bhutanese teachers reported high self-efficacy about teaching the Bhutanese Grade 10 mathematics curriculum. Also, as claimed by Caprara et al., (2003), the teachers' high self-efficacy might have contributed to the majority of teachers being content with their job, in that as many as four in five were satisfied with their job.

The efficiency of homework assigned to students is known to depend on frequency, duration, and purpose of the homework. The findings from this study showed that the homework was most effective, as measured in terms of student achievement, when it is assigned to students for almost every lesson, with every homework occasion requiring a completion time ranging from half to one hour. In addition, the efficiency of homework also peaked when students were engaged in correcting their homework with support from their teachers. As reported in the literature (Brock, et al., 2007; Cooper, 1989; Cooper, et al., 1998; Cooper, et al., 2006; Muhlenbruck, et al., 2000; Mullis, et al., 2008; OECD, 2007a), these findings suggest that teachers need to determine the best frequency, duration, and monitoring strategy for homework to support students' learning.

It is in the culture of every school to conduct tests or examinations for different purposes, such as monitoring students' learning progress and deciding students' promotion, but schools differ in the frequency with which the tests are conducted and the types of questions used in the tests. This study showed that the maximum student performance score corresponded to once-a-month frequency of tests. The correspondence of the once-a-month frequency of the test to the maximum test score suggests that teachers get sufficient time to evaluate the answer scripts and provide feedback to students, and that the students get sufficient time to prepare for the tests within this frequency. Further, the study showed that approximately one in five teachers reported using questions that required higher-order thinking skills in almost every test. Conversely, approximately four in five teachers reported using questions that required students to use procedural knowledge and skills in almost every test. This finding suggests that teachers either set low standards for their students or that their teaching focuses more on procedural knowledge and skills than on engaging students in solving mathematical problems requiring the use of higher-order thinking skills. Because generally higher test standards lead to greater student effort (Phelps, 2005), the possibility of teachers setting low test standards has implications for students, teachers, schools, and policymakers as reviewed in Section 2.5.1 of Chapter 2. The Ministry of Education may consider developing strategies to monitor the standard of tests or examinations conducted by schools with the aim of supporting them in providing higher quality tests to students.

The use of calculators and computers in the mathematics class has been known to free students from performing routine calculations, provide students with time for deep learning, and enable students to visualise mathematical concepts, resulting in better student outcomes (Ellington, 2003; Ruthven, et al., 2004). This study showed that Bhutanese Grade 10 mathematics teachers used calculators in the class for checking answers, doing routine computations, solving complex problems, and exploring number concepts. However, not all Bhutanese Grade 10 students had access to calculators in the mathematics class. Approximately, 65% of students were taught by the teachers who reported that their students did not have access to calculators in the mathematics class. Given that calculators are used in the class, and that they are known to associate with student outcomes (Ellington, 2003; Ruthven, et al., 2004), students without access to calculators may be significantly disadvantaged in learning mathematics. Therefore, schools may explore ways to help all students get the benefits from using calculators in the mathematics class. The use of computers in the mathematics class was minimal; on average, 74% of students were taught by teachers who reported never using computers in the mathematics class. Because computers can help students discover mathematics principles and concepts, practise skills and procedures, look up ideas and information, and process and analyse data (Mullis, et al., 2008; Ruthven, et al., 2004), the Ministry of Education may explore ways to increase the use of computers in the mathematics class.

10.7 Emerging Themes from Analyses of the School Questionnaire (Chapter 8)

Schools' capacity to provide an enabling space for teachers and students to engage in effective teaching and meaningful learning has been known to depend on a range of factors. The factors such as school policies and practices, school climate, school leadership, and school resources have been receiving increasing attention in international educational assessments (Mullis, et al., 2008; OECD, 2007a). Knowledge about these factors has the potential to help stakeholders in the Bhutanese education system to make Bhutanese schools more effective. Overall, this study showed that there is scope for the Bhutanese education system to standardise school policies and practices, improve school climate, develop school leadership capacity, and invest in school resources.

School policies and practices related to admittance, assessment, ability grouping, parental involvement, and autonomy have diverse implications for schools. School admittance policies facilitate student mobility between schools. As brighter students and good teachers usually prefer high-performing schools to low-performing schools, high-performing schools may benefit from brighter students and less teacher turnover (OECD, 2007a). Conversely, low-performing schools may suffer from losing brighter students and more teacher turnover. This study showed that the proximity of students' residence to school and directives from the Ministry of Education were the dominant criteria for school admittance. This suggests low student mobility between schools and little parental choice. In other words, students have little option to enrol in a school of their choice, and parents have little option to make schools accountable for their children's educational attainment level. Because student mobility is associated with student achievement (Cullen, et al., 2005; Gibbons, et al., 2008; Lavy, 2010; OECD, 2007a; Soderstrom & Uusitalo, 2010), ways to facilitate it in Bhutanese schools could be explored. Despite little pressure from students and parents because of low student mobility, a range of assessment practices was implemented in Bhutanese schools, resulting in the establishment of accountability. Similar to the findings reported in the literature (Cizek, 2005; Clark, Lee, et al., 2008; OECD, 2007a; Phelps, 2005; Stecher, 2002), assessment results were used for monitoring student learning, deciding student promotion, appraising teacher performance, and connecting with parents. Further, schools reported a range of expectations for parents to engage actively in their children's education. Although, it is beyond the reach of this study to confirm the extent to which those expectations or opportunities helped parents involve actively in school activities, the roles that schools expected parents to play in their children's learning are in line with the findings reported in the literature (Bowen & Lee, 2006; Driessen, et al., 2005; Fan & Chen, 2001; Grolnick, et al., 1997; Grolnick & Slowiaczek, 1994; Izzo, et al., 1999; Jeynes, 2003, 2005, 2007; Simon, 2001, 2004; Yan & Lin, 2005). This finding also suggests that school principals are aware of the benefits of engaging parents in their children's education, signalling the potential for greater collaboration between schools and parents for their children's benefits. The final dimension of school policies and practices that this study examined was school autonomy in the following areas: formulating and allocating school budget; establishing school disciplinary policies; establishing school assessment policies; school staffing; and selecting textbooks. This study showed that schools differed in their perceptions about autonomy in these areas. Because school autonomy has been known to relate to student achievement (Fuchs & Wobmann, 2007; Maslowski, et al., 2007; OECD, 2010; West, et al., 2010; Wobmann, 2007), different perceptions among schools about their autonomy may have implications for student outcomes. Therefore, the Ministry of Education can provide schools with explicit guidelines on their autonomy, with the aim of establishing consistency in the exercise of autonomy by schools.

An enabling school climate, including an orderly respectful atmosphere, could lead to an improved achievement orientation (Brand, et al., 2003; Marzano, 2001; Scheerens, 1990, 1997, 2000; Scheerens, et al., 2003). Therefore, school principals' perceptions about school climate, as construed in terms of student behaviour, teacher behaviour, teacher consensus, and teacher morale, could provide stakeholders in the Bhutanese education system with knowledge and opportunities to foster positive school climate. Similar benefits are also reported in the OECD (2004a). This study showed that there is much opportunity for Bhutanese schools to improve their climate. Some of the potential areas for improvement are: reducing student absenteeism and truancy; increasing teachers' capacity to cater for individual students' learning needs; raising awareness among teachers about the benefits of having high expectations of students; providing teachers with the opportunities for collaboration with one another to foster consensus; and engaging teachers in school activities to raise their morale and enthusiasm. Because unfavourable school climate has been associated with student disengagement (Finn & Voelkl, 1993), student misconduct, aggression, and behavioural problems

(Wilson, 2004); and drug abuse and delinquent behaviour (Battistich & Hom, 1997; Battistich, et al., 1995), the implications from an unfavourable school climate for quality education are clearly apparent.

Educational leadership has been widely viewed as an influential factor of school effectiveness. Among different educational leadership models reported in the literature (Hallinger & Heck, 1996; Leithwood & Jantzi, 1990), this study showed that the principals of Bhutanese schools practised instructional leadership. While the principals' self-reported practising different dimensions of instructional leadership reported in the literature (Hallinger, 1989, 1990, 1994, 2003, 2005; Hallinger & Heck, 1996, 1998), the principals seemed to be less proficient in instructional leadership. Principals appeared to be setting goals more frequently than performing any other leadership roles in contrast to the usual habit of setting goals once at the beginning of the school year, and possibly followed by a few reviews in the course of the school year. This suggests that principals changed school plans more often than implementing them. Because the success of instructional leadership has been known to depend on the proficiency of the principals in its concept and implementation strategies (Hallinger & Heck, 1996, 1998; Maeyer, et al., 2007; Marks & Printy, 2003; Witziers, et al., 2003), it appears that Bhutanese school principals need more support for, and opportunities to, develop their leadership capacity.

School resources are malleable (Scheerens, et al., 2003), and because teaching and learning in schools have been known to relate to school resources (Greenwald, et al., 1996; Hanushek, 1997; Hedges, et al., 1994; Marzano, 2001; Scheerens, 2000; Wayne & Youngs, 2003; Wenglinsky, 2002), knowledge about school resources has the potential to guide stakeholders in the Bhutanese education system in developing and deploying scarce educational resources to where they are most needed. This study showed that Bhutanese schools were constrained by the lack of both human and material resources. For instance, more than one-half of Bhutanese Grade 10 students were in schools whose principals reported that their schools capacity to provide effective instruction was hindered by the lack of educational resources. At the surface level, this finding shows that Bhutanese schools need to be supplied with adequate educational resources. However, as reported in the literature (Greenwald, et al., 1996; Hanushek, 1997; Hedges, et al., 1994; Marzano, 2001; Scheerens, 2000; Wayne & Youngs, 2003; Wenglinsky, 2002), the resource-related challenges faced by schools may also stem from improper deployment of the resources available within schools. Therefore, the Ministry of Education may develop educational interventions aimed at building the capacity of its schools in effective use of educational resources.

10.8 Emerging Themes from Analyses of the Focus Group Interview (Chapter 9)

One of the assumptions of the proposed national educational assessment model has been that its context level provides enabling conditions for other levels. The variables commonly identified for the context level are external achievement stimuli, school environments, and evaluation of national educational policies (Creemers & Kyriakides, 2008; Scheerens, 1992). This study assumed that these variables play a vital role in raising the standard of the Bhutanese education system.

School performance is subject to a range of stimuli from external agents. Parents can influence schools to perform better by being involved in school activities, educational consumers can influence schools to perform better by choosing high-performing schools, and governments can influence schools by enacting enabling policies in diverse areas of school functioning (Scheerens, et al., 2003). In line with the literature, the focus group endorsed the need for schools to connect with parents (Fan & Chen, 2001; Hoover-Dempsey, et al., 2001; Hoover-Dempsey & Sandler, 1995, 1997; Jeynes, 2007; Ritblatt, et al., 2002; Yan & Lin, 2005). However, the focus group pointed out challenges in reaching out to parents because of the lack of a common communication platform between school and parents. The medium of communication such as school magazines, brochures, calendars, student performance reports, and media were viewed as ineffective because the majority of parents might not be able to absorb them due to the low literacy rate of the country. Therefore, corroborating the findings in Ritblatt, et al. (2002), the lack of proper communication seems to be a key barrier to parental involvement in the activities of Bhutanese schools. This perception also raises doubts about parents' ability to assist their children with homework and other academic activities, which is a challenge being faced by parents with low SES and less formal education (Epstein, 1995; Epstein & Sanders, 2006). However, because the parental involvement in school activities or children's learning spans beyond academic activities in the forms of parenting, volunteering, decision-making, and collaboration with communities (Epstein & Sanders, 2006), alternative school outreach programmes should be able to involve parents in school activities.

The analyses of the focus group interview also showed that school-by-school publication of student achievement results, post-school enrolments in further studies, and scholarship were being viewed as a powerful means of enabling schools to perform better. However, in line with cautions raised in Goldstein and Leckie (2008), a fear of negative practices associated with school "league tables" is also prominent in the focus group interview. This finding suggests indecision among stakeholders in using school league tables to drive school performance. The findings that there were challenges in communication between parents and schools, and that school league tables were viewed more as a nuisance than a blessing, suggest that parents do not have much access to information about schools, making it difficult for them to choose high-performing schools for their children. As expected, the analyses of the focus group interview showed that school choice is almost non-existent in Bhutan because schools are allocated to students by the Government: student mobility is perceived as the prerogative of the Government. Similar to the findings reported in the literature, a concern about students migrating to high-performing schools from low-performing schools when parents and students have the freedom of choosing schools was prevalent among stakeholders in the Bhutanese education system (Gibbons, et al., 2008; Gibbons & Telhaj, 2007; Hanushek, et al., 2004). The absence of school choice for parents in the Bhutanese education system also suggests little autonomy of schools in admitting students.

Besides little autonomy of schools in the admission of students, schools also had little autonomy in procuring teaching and learning materials, recruiting and terminating teachers, and using school budgets. In principle, however, a general perception among the focus group was that the autonomy of schools in these areas would make the schools directly responsible for their performance. This perception is similar to the findings reported in the literature where school autonomy has been known to improve school accountability (Fuchs & Wobmann, 2007; Maslowski, et al., 2007; OECD, 2010; Timar & Roza, 2010; West, et al., 2010; Wobmann, 2007). One persistent reason that emerged from the analyses of the focus group interview for not granting autonomy to schools in these areas was a perception that the schools were not competent enough to discharge the responsibilities associated with the areas of autonomy. As suggested by Grauwe (2005), policymakers may grant autonomy to schools in conjunction with supportive strategies to develop capacities of principals, teachers, and communities with a clear focus on autonomy as a tool for improving school performance and establishing equity in the school education system. The analyses of the focus group interview data also revealed that the prospect of using performance-based funding or incentives as a means to improve school accountability is unlikely. It emerged from the focus group that the Bhutanese education system did not have the capability to develop a comprehensive list of school performance indicators other than the examination-oriented ones, making performance-based funding or incentives irrelevant. However, some aspects of performance-based funding or incentives, as reported in Foster and Marquart (1984), such as career ladder, meritorious promotion, master teacher, and in-service training are already in practice in Bhutanese schools, and yet these aspects of performance-based funding are not prominent in the perception of the focus group about performance-based funding. A possible implication of this finding is that teachers in the Bhutanese education system progress in their career and get scholarships for further studies based on a seniority attained because of the number of years, not because of outstanding performance.

Much as the institution of autonomy and accountability facilitates successful teaching and learning in schools, enabling learning environments in schools are also fundamentally important for successful teaching and learning experiences in schools. The focus group highlighted the importance of physical infrastructural resources for developing enabling learning environments in schools. The focus group overlooked the intangible constructs (e.g., school climate, school leadership, and student engagement) in its list of the attributes of enabling learning environments in schools. At its best, this finding suggests that the focus group took for granted the intangible aspects of enabling school climate. Less favourably, this finding may be indicative of the lack of a comprehensive knowledge of enabling school environments among stakeholders in the Bhutanese education system. The focus group's perspective of the relation between enabling learning environments and student achievement suggests that adequate physical infrastructural resources would develop enabling learning environments in schools and eventually result in child-oriented education. However, the literature shows that physical infrastructural resources and other intangible attributes of enabling learning environments are not always compensatory, but are rather disjunctive (Barber & Mourshed, 2007; Hanushek, 1997; Hoy & Hannum, 1997; OECD, 2004a; Wilson, 2004). Therefore, an equal emphasis on physical infrastructural resources and other intangible attributes might be made to develop enabling learning environments in schools.

National educational policies are wide-ranging, and it is not the intention of this study to analyse them other than finding out if there existed a culture of policy analysis in the Bhutanese education system. The focus group findings suggested that educational policies are rarely analysed and evaluated. As stated in Bridgman and Davis (2007), not evaluating the policies would result in the lack of knowledge about: how the policies have achieved their objectives; who to hold accountable for the implementation of the policies; and the directions and clues for future policy making. These implications are serious because they imply the lack of policy analyses and the loss of opportunities for policy learning (Bridgman & Davis, 2007; Parsons, 1995).

10.9 Interplay of the Findings

The assumptions underlying the national educational assessment model presented in Section 3.5.1 of Chapter 3 provides the bases for analysing the interactions of the findings. One of the assumptions relevant to such analyses is that the higher levels of the model provide an enabling environment for the lower levels. This section draws together the findings at different levels of the model and discusses their relations with one another in the light of this assumption.

The analyses of data at the context level of the model revealed that the Bhutanese education system has much scope for reaping the benefits of introducing educational interventions such as external achievement stimuli, effective school environments, and evaluation of national education policies. External achievement stimuli are usually designed to make schools accountable to stakeholders (Scheerens, 1992, 2001). It emerged from the focus group interview that Bhutanese schools are well shielded by the Ministry of Education from being accountable to other stakeholders. The fact that Bhutanese schools are less exposed to the workings of external achievement stimuli may make them less accountable to stakeholders other than the Ministry of Education or its agents. The other implication from this finding is that schools may not receive much support from parents and community because they play little role in school accountability. The Ministry of Education also views physical school resources as the factors of enabling learning environments with little regard for other factors. Because schools are accountable to the Ministry of Education, less emphasis by the Ministry on other factors such as school leadership, school climate, and student engagement may make schools complacent about these factors. Finally, the non-existence of the practice of policy analysis in the Ministry of Education implies that lessons are not learnt from policy successes or failures. Either of these scenarios shows the lack of systemic accountability and scientific approaches to policy development (Bridgman & Davis, 2007).

Bhutanese schools are directly accountable to the Ministry or its agents. Given little systemic accountability and research-based approaches to policy development in the Ministry of Education, it is likely that schools too have similar problems. The analyses of school questionnaire data showed that school policies and practices need standardization, that school climate needs greater emphasis, that school leadership needs further support, and that school resources need strategic deployment modalities. The lack of standards for school policies and practices in the areas of student admittance, student ability grouping, parental involvement, and school autonomy may pose challenges to the Ministry of Education in monitoring and evaluating its schools. Because schools are well shielded by the Ministry of Education from public accountability, the lack of a standardised monitoring and evaluation system may give schools unwarranted licence for their low performance. In addition, schools may not be able to benefit from professional feedback and interventions that usually ensue with monitoring and evaluation.

Dissimilar policies and practices and modest autonomy of schools may also determine the extent to which teacher-related factors interact with student achievement. This study showed that teachers' demographic profile, professional collaboration, professional development, appraisal and feedback, school climate, beliefs about teaching approaches, use of teaching approaches, teaching constraints, self-efficacy, homework, use of tests and examinations, and use of calculators and computers played a significant role in student achievement. The study also showed a number of ways that these factors could be best used in the interest of schools. Gender parity, age and experience, and qualifications showed up in the teacher demographic profiles as the areas requiring emphasis in teacher deployment policies. Teachers' needs for opportunities and support for participation in professional development emerged as the cause for the low rate of teacher participation in professional development. The teacher appraisal system in schools needs improvement. School climate and its significance in enabling positive environments need greater emphasis in schools. Teaching paradigms and their corresponding teaching strategies need to be made explicit to teachers. Frequency and duration of homework need to be included in schools' homework policies. Schools need to encourage teachers to test students with test items requiring the demonstration of higher-order thinking skills. Schools need to work out strategies to provide uniform student access to calculators and computers. In sum, the need for schools to improve in various teacher-related factors suggest that schools need regular professional feedback from external monitoring and evaluation agents and equivalent support in assimilating the feedback in their overall functioning.

It is likely for an education system that lacks a systemic assessment and evaluation mechanism to overlook important student-related factors when developing interventions to improve student achievement. This study showed up a number of student-related factors that the Ministry of Education could consider for potential interventions. Students' gender, age, socio-economic status, engagement, motivation, self-efficacy, and self-regulated learning skills have scope for interventions. Similarly, homework, ICT, classroom management, and school climate have capacity for interventions. Given that students performed better when the desirable aspects of these factors were reported, interventions aimed at developing these aspects have the potential to improve student achievement. Therefore, the Ministry of Education may pursue independent research in these student-related factors, with the aim of developing interventions to improve them.

In summary, drawing on the model developed in this study, there is great scope for interventions across different levels of the Bhutanese education system and a great potential for rapid improvement in the quality of the Bhutanese education system because of the implementation of these interventions. This scope is further emphasised in the next chapter in the light of a range of conclusions from the study.

Chapter 11
CONCLUSIONS

11.1 Introduction

This chapter first presents conclusions from the study. These are followed by sections on the significance and limitations of the study, the recommendations from the study, and the directions for future research.

11.2 Conclusions from the Study

This study sought to develop a national educational assessment model for Bhutan with the capacity to address ten outcomes presented in Section 1.3 of Chapter 1. Accordingly, a national educational assessment model for Bhutan has been presented in Chapter 3, and the expected capacity of the model has been demonstrated empirically in Chapters 5 through 9 within the scope of the ten outcomes. As the outcomes were the key indicators of the success of the thesis, this section will present a brief summary of how this study has met each outcome.

Outcome 1: Knowledge profiles of students about their comprehension of school curricular content

Chapter 5 demonstrated that Bhutanese students did not fully attain curricular goals. Students need to be provided with greater opportunities for reading mathematical information, interpreting mathematical formulae, making discretional use of mathematical procedures and arguments, and drawing logical conclusions from mathematical calculations and data analyses.

Outcome 2: Knowledge about the factors related to effective schooling and their effects on student outcomes

Drawing on literature, Chapter 2 presented a range of student-, teacher-, school-, and context-related educational effectiveness factors. Student-related factors included the following: demographic profile; homework; motivation; self-beliefs; self-regulation; learning preferences; ICT; classroom management; and school climate. Teacher-related factors consisted of the following: demographic profile; professional development; appraisal and feedback; self-efficacy; school climate; classroom management; effective teaching components; homework; and assessments. School-related factors included school policies, and practices: school climate; educational leadership; and school resources. Context-related factors included the following: national or regional policy for education; evaluation of educational policy; educational environment; achievement stimulants from higher achievement levels; development of educational consumerism; and school category (e.g., rural, urban, private, government).

Chapters 6 through 9 demonstrated that most of the above factors related to the Bhutanese education system and its students, teachers, and schools. Student, teacher, and principal responses that demonstrated strong alignment with preferable practices identified in the literature performed better in educational outcomes than students, teachers, and schools where there was not this alignment.

Outcome 3: Knowledge about student outcomes at school, district, region, and national levels

Chapter 5 showed that there is much scope for the Bhutanese education system to benefit from cross-fertilization of ideas among schools because some schools were more effective than others. This calls for the Ministry of Education or School Districts to facilitate exchange of ideas between schools.

Outcome 4: Knowledge about the preparedness of Bhutanese students to meet the challenges of the future

Chapter 6 provided insights into how well Bhutanese students are prepared for the challenges of the future as judged in terms of their preparedness to pursue life-long education. Students tended to be more instrumentally motivated, more performance goal-oriented, more proficient in lower-order thinking skills, and poorly equipped with ICT knowledge and skills. Therefore, Bhutanese students needed to develop more fully their readiness for life-long learning, and (by implication) their capacity for future challenges.

Outcome 5: Knowledge about the skills that Bhutanese students need to adapt to rapid societal and technological change.

Chapter 5 showed that Bhutanese students are more competent at solving mathematical problems requiring lower-order thinking skills than solving mathematical problems requiring higher-order thinking skills. Because higher-order thinking skills are vitally important for life-long learning, which in itself is considered as a means to cope with rapid societal and technological change, one of the implications is that Bhutanese students need to be provided with greater learning opportunities related to higher-order thinking skills.

Outcome 6: Knowledge about teachers and teaching and their effects on student outcomes

Chapter 7 covered a range of teacher-related factors that needs interventions to maximise their contributions to student achievement. From the point of teachers’ demographic profiles, teacher deployment policies need to include teachers’ age and experience, and the number of female teachers in the mathematics teacher workforce needs to be increased. Teachers' professional collaborations indicate that they need to be guided by research findings when they collaborate and interact with one another on the matters related to teaching and learning. Further, teachers need more needs-based opportunities to participate in professional development. From the point of teacher appraisal, both teachers and principals need training in the efficient use of appraisal instruments. From the point of learning environments, greater attention needs to be accorded to the benefits of, and the ways to improve, classroom and school climates. From the view of teachers' pedagogical knowledge and skills, teachers need to be provided with situations that call for them to demonstrate a repertoire of teaching strategies and their ability to map those strategies to relevant teaching paradigms in response to the diverse learning needs of students. With reference to teachers' view of themselves, teachers need opportunities to engage in events that give them experiences to raise their self-efficacy and make them aware of the positive relation between high self-efficacy and the quality of their teaching. In regard to homework, teachers need to consider the frequency, the duration, and the feedback aspects of homework when planning homework for students. From the point of textbooks, teachers need to use alternative teaching and learning resources in proportion to their use of prescribed textbooks to broaden their choice of teaching and learning resources. From the point of assessment, teachers need to assess students in the content areas that require more higher-order thinking skills than assessing those that require low-order thinking skills. Finally, teachers need to explore all possibilities of according a uniform access for students to calculators and computers in the mathematics class.

Outcome 7: Knowledge about the educational structures and practices that maximise the learning opportunities

Chapter 8 showed that schools have a lot to learn from the strengths and weaknesses of one another in terms of their structures and practices. Learning can occur in the areas of school policies and practices, school leadership, school climate, and school resources. Such support may be in the form of professional development or inter-school visits. From the point of school climate, schools need to engage students and reduce student absenteeism and truancy by enabling their teachers to meet individual students' learning needs and encouraging them to demonstrate high expectations of their students. In addition, schools need to provide opportunities for teachers to develop consensus in their views related to teaching paradigms and pedagogies and their expectations from students. From the point of school policies and practices, school admittance policy needs to be reviewed in terms of its benefits to students, parents, and schools. Not many schools used ability grouping, and schools in which it is used need to review their policies and practices in the light of the benefits to student learning.

Schools differed in their reports of the areas in which they reported to have autonomy, indicating the lack of uniformity among the schools in their knowledge of autonomy from the Ministry of Education or the School Districts. Schools need to be clear about the extent of autonomy they have from the Ministry of Education or the School Districts. Schools also need support for delivering effective instructional leadership.

Chapter 9 showed interesting insights into the perspectives of the policymakers in the Bhutanese education system. The policy makers endorsed the importance of recognising student achievement through publication and other meritorious awards. Policies need to be formulated on making schools more accountable to the stakeholders through granting greater autonomy to the schools. Factors other than resources may also be given priorities in the policies related to effective learning environments. A culture of policy evaluations needs to be developed in the Ministry of Education.

Outcome 8: Knowledge about equity in, and accessibility to, educational resources

Chapter 6 showed that inequitable access to educational resources and learning opportunities is an issue in the Bhutanese education system as indicated by the difference in the achievement of students from high and low SES family backgrounds. The Ministry of Education has great opportunity to use its schools as agents for reducing the gap between low and high SES families by using SES as a factor in its policies on equitable distribution of educational resources to schools. In addition, as shown in Chapter 8, schools need adequate supply of resources for effective teaching and learning together with support for efficient use of the resources. Overall, the fact that many student-related factors are associated with student achievement in addition to SES suggests that it may be possible to offset the negative effects of inadequate educational resources by the positive effects of other student-related factors.

Outcome 9: Knowledge about the standard of the Bhutanese education system as compared to the standard of other countries

Chapter 5 showed the international benchmark of the Bhutanese education system as being lower than the OECD’s average, and implied that there is great potential for Bhutan to benefit from interacting with both high- and low- performing PISA countries.

Outcome 10: A database for studying student achievement over time.

This outcome relates to data used in this study and the database created, and these will be discussed in the next section.

11.3 Significance and Limitations of the Study

The significance of this study lies in its outcomes and the methods used for generating them. Apart from the nine outcomes and their significance mentioned earlier, this study also produced a database for future research. The database, developed from data used in this study, will be able to provide future researchers with data on a range of factors related to teaching and learning in the context of the Bhutanese education system. For instance, individual educational effectiveness factors from the database merit separate discussions. Furthermore, the literature review and the resulting educational effectiveness factors from the review will provide educators in Bhutan with comprehensive information about the latest research findings and the current debates in education that have the potential to further their existing knowledge and motivate them to participate in the debates among the communities of educators.

The expected outcomes of this study have set the research primarily in the Bhutanese context, but the methods and procedures used in this study have relevance to assessment procedures of other like nations. First, the method of linking the test items from PISA to the national curriculum has the potential to close the on-going debate about the relevance or irrelevance of PISA to the national curriculum of the countries that participate in PISA. Second, the method of benchmarking a country's education system with the education systems of the countries that participate in PISA, without necessarily participating in PISA, has the potential to enable countries like Bhutan to benefit from PISA as demonstrated in this study. Third, assessing a nation's education system by incorporating students, teachers, schools, and Ministry of Education provides stakeholders with comprehensive knowledge about the health of their education system. Such comprehensive knowledge has the potential to guide stakeholders in developing interventions for improving the quality of their education system. In sum, the methods used in this study have the capacity to generate comprehensive knowledge about a country's education system with implications for developing strategic educational interventions as demonstrated.

Like any research study, this study too has its limitations. First, the study is cross-sectional; that is, it is conducted as a one-off PhD study. Therefore, the study cannot provide cause and effect interpretations between, and inferences from, educational effectiveness factors. Second, the questionnaires used in the study are self-administered by the participants. Like any self-reported data, the data from the survey questionnaires are subjective, and are different from objectively measured data. In spite of all necessary measures to safeguard the study against possible concerns about its validity from strategic response behaviours by all participants, it is still possible that such concerns have permeated into the study. Some participants might have responded to the questionnaires to comply with what they perceived to be socially desirable at the cost of the actual responses;

therefore, the subjectivity of the self-reported data needs to be borne in mind when drawing inferences from the study. Third, the need to include diverse range of educational effectiveness factors and the paucity of space for sufficient account of the implications from data on the factors limited the study to engaging only in summary discourses of its findings. Because of the lack of exhaustive discourses, which are commonly found in doctoral theses with single, explicit research problems, this study has presented its findings as indicators for further research in the context of the Bhutanese education system. Finally, the lack of an adequate fund for field research limited the number of participants for the qualitative part of the study to only one focus group. This constraint deprived the study of the broader perspectives on the research questions discussed in its qualitative section.

In summary, this study developed a national educational assessment model for the Bhutanese education system, and demonstrated the potential of the model to generate diverse knowledge about educational effectiveness factors relating to the success of the Bhutanese education system as measured in terms of student achievement on the Mathematics test. The outcomes from the model indicated much scope for interventions in the Bhutanese education system at student, teacher, school, and national levels.

11.4 Recommendations

This study aimed to develop a national educational assessment model to collect and analyse data about the performance of the Bhutanese education system in relation to Mathematics. However, the real utility of the study will depend on how its findings are assimilated into the Bhutanese education system. A matrix of follow-up actions and agents responsible for the actions to facilitate systemic assimilation of its findings remains a desirable element of this study, which can only be generated in consensus with stakeholders in the Bhutanese education system.

With the matrix in place, a series of pilot interventions at student, teacher, school, and national levels can be initiated to improve the quality of the Bhutanese education system. At the student level, teachers or schools may initiate interventions in student motivation, self-beliefs, self-regulated learning skills, and ICT, with the aim of making students more ready for life-long learning. At the teacher level, schools or the Ministry of Education may initiate interventions in professional development, teaching paradigms and pedagogies, and teacher appraisal and feedback to enhance teaching competency of teachers. At the school level, the Ministry of Education may initiate interventions in school leadership, school climate, school resources, and school policies and practices to improve school effectiveness. At the national level, the Ministry of Education may initiate interventions in school autonomy, school accountability, school choice, school environments, and policy evaluation.

In addition to their relevance to teachers, schools, and the Ministry of Education, the findings from the study may also be of interest to other stakeholders in the Bhutanese education system. Teacher training colleges may find most of the findings reported about student, teacher, and school characteristics as first-hand, contextualized knowledge worthy of sharing with student teachers. In some cases, the findings may also assist the teacher training colleges in prioritising teaching areas or in identifying problems for potential research. Other post-school colleges and institutions may find students' knowledge profiles interesting as they provide insight into what students know. Such insights may help the colleges and the institutions to determine the readiness of potential students for their courses. Parents, public, and educational institutions may find the international benchmarks interesting as they compare Bhutanese students with students from other countries.

11.5 Directions for Future Research

In spite of persistent claims about the capacity of the study to contribute to the improvements in the Bhutanese education system, future research beyond this study can greatly enhance its current value. This future research could include the involvement of other key school subjects, refining and further developing the research methods, conducting impact studies on the interventions, and establishing the study as a periodic assessment programme.

It would be possible to adapt this study to include key school subjects such as science and English, in addition to mathematics, in line with other well-established international educational assessment programmes. Such modifications would greatly increase the capacity of the study to provide holistic knowledge about the current state of the Bhutanese education system. Further, to achieve full alignment between the test and curriculum standards, increasing the number of test items without increasing the testing time, in line with the purpose of the Balanced Incomplete Block Design method of sampling test items, would be a worthwhile modification to the approaches followed in developing the Mathematics test used in this study.

Application of multi-level analysis techniques and generation of trends in student achievement would be desirable improvements in the overall design of this study. Methods to quantify data collected at the context level of the model may be explored so that it becomes possible to obtain the total variance in student achievement explained by the model. This would also make it possible for researchers to calculate the variances in student achievement explained by different levels of the model. Further, it would be possible to determine interaction effects of educational effectiveness factors, and calculate the amount of variance in student achievement explained by individual educational effectiveness factors. Such knowledge would assist stakeholders in grading the importance of the educational effectiveness factors in terms of the strength of their relationship with student achievement. Another area for future study would be to study changes in student achievement over time, which is a prominent design element in international large-scale assessments. A variant of this study could be conducted every three years in the Bhutanese education system to study trends in student achievement and in international benchmarking. In addition, other countries who are not involved in large-scale assessment programmes could adapt this approach.

As is the case with any interventions, implementation of each of the interventions suggested in this study would need to be followed by an impact study. Case studies or quasi-experimental studies may be pursued to assess the impact of the interventions on the Bhutanese education system.

Overall, an emerging theme from the study indicates that teaching and learning are most effective when student, teacher, school, and contextual characteristics in a nation's education system complement one another. This theme underscores the importance of considering these characteristics when developing and sustaining national educational policies and interventions.

References

Abedi, J., Lord, C., & Hofstetter, C. (1998). *Impact of Selected Background Variables on Students' NAEP Math Performance*. Los Angeles: University of California.

Adams, R. J. (2003). Response to 'Cautions on OECD's Recent Educational Survey (PISA)'. *Oxford Review of Education, 29*(3), 377-389.

Adams, R. J., & Wilson, M. (1996). Formulating the Rasch Model as a Mixed Coefficients Multinomial Logit. In G. Engelhard & M. Wilson (Eds.), *Measurement: Theory into Practice* (Vol. 3, pp. 143-166). Norwood, NJ: Ablex Publishing Corporation.

AERA, APA, & NCME. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational research Association.

Aiken, L. R. (1997). *Psychological Testing and Assessment* (9 ed.). Needham Heights, MA: Allyn and Bacon.

Aiken, L. R., & Growth-Marnat, G. (2006). *Psychological Testing and Assessment* (12 ed.). Boston: Pearson Education Group, Inc.

Akin-Little, K. A., Little, S. G., & Laniti, M. (2007). Teachers Use of Classroom Management Procedures in the United States and Greece: A Cross-Cultural Comparison. *School Psychology International, 28*(1), 53-62.

Alsup, J. (2004). A Comparison of Constructivist and Traditional Instruction in Mathematics. *Educational Research Quarterly, 4*(28), 3-17.

Ames, C. (1992). Classrooms: Goals, Structures, and Student Motivation. *Journal of Educational Psychology, 84*(3), 261-271.

Ames, C., & Archer, J. (1988). Achievement Goals in the Classroom: Students' Learning Strategies and Motivation Processes. *Journal of Educational Psychology, 80*(3), 260-267.

Anderson, C. S. (1982). The Search For School Climate: A Review of the Research. *Review of Educational Research, 52*(3), 368-420.

Anderson, J. C., & Gerbing, D. W. (1988). Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin, 103*(3), 411-423.

Archer, L., Halsall, A., & Hollingworth, S. (2007). Class, gender, (hetero) sexuality and schooling: paradoxes within working-class girls' engagement with education and post 16-aspirations. *British Journal of Sociology of Education, 28*(2), 165-180.

Archived: GOALS 2000: Educate America Act. from http://www.ed.gov/legislation/GOALS2000/TheAct/index.html

Arthaud-Day, M. L., Rode, J. C., Mooney, C. H., & Near, J. P. (2005). The Subjective Well-Being Construct: A Test of its Convergent, Discriminant, and Factorial Validity. *Social Indicators Research, 74*, 445-476.

Ashcraft, M. H. (2002). Math Anxiety: Personal, Educational, and Cognitive Consequences. *Current Directions in Psychological Science, 11*(5), 181-185.

Ashcraft, M. H., & Moore, A. M. (2009). Mathematics Anxiety and the Affective Drop in Performance. *Journal of Psychoeducational Assessment, 27*(3), 197-205.

Babbie, E. (1990). *Survey Research Methods* (2 ed.). California: Wadsworth, Inc.

Bagozzi, R. P., & Phillips, L. W. (1982). Representing and Testing Organizational Theories: A Holistic Construal. *Administrative Science Quarterly, 27*(3), 459-489.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing Construct Validity in Organizational Research. *Administrative Science Quarterly, 36*(3), 421-458.

Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Bhavioural Change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Printice-Hall.

Bandura, A., & Schunk, D. H. (1981). Cultivating Comptence, Self-Efficacy, and Intrinsic Interest Through Proximal Self-Motivation. *Journal of Personality and social Psychology, 41*(3), 586-598.

Barber, M., & Mourshed, M. (2007). How the world's best-performing school systems come out on top. Retrieved 20 February 2007, from http://www.mckinsey.com/clientservice/socialsector/resources/pdf/Worlds_School_systems_final.pdf

Barton, P. E. (2002). *Perspectives on Background Questions in The National Assessment of Educational Progress*. Washington, D.C: National Assessment Governing Board.

Bathory, Z. (1992). Hungarian experiences in international student achievement surveys. *Prospects, 22*(4), 433-440.

Battistich, V., & Hom, A. (1997). The relationship between Students' Sense of Their School as a Community and Their Involvement in Problem Behaviours. *American Journal of Public Health, 87*(13), 1997-2001.

Battistich, V., Solomon, D., Kim, D.-i., Watson, M., & Schaps, E. (1995). Schools as Communities, Poverty Levels of Student Populations, and Students' Attitudes, motives, and Performance: A Multilevel Analysis. *American Educational Research Journal, 32*(3), 627-658.

Beaton, A. E., & Gonzalez, E. (1995). *NAEP Primer*. Chestnut Hill, MA: Boston College.

Bell, B. S., & Kozlowski, S. W. J. (2002). Goal Orientation and Ability: Interactive Effects on Self-Efficacy, Performance, and Knowledge. *Journal of Applied Psychology, 87*(3), 497-505.

Betts, J. R., & Shkolnik, J. L. (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review, 19*, 1-15.

Bhutan Board of Examinations. (2004). *National Education Assessment in Bhutan: A Benchmark Study of Student Achievement in Literacy and Numeracy at Class VI, 2003*. Thimphu.

Bhutan Board of Examinations. (2008). *NEA 2006 National Education Assessment, Class 10 Mathematics and English Report: Findings from the National Education Assessment of Mathematics and English of class 10 in the Schools of Bhutan*. Thimphu: Bhutan Board of Examinations.

Black, T. R. (1999). *Doing Quantitative Research in the Social Sciences: An Integrated Approach to Research Design, Measurement and Statistics*. London: SAGE Publications Ltd.

Boekaerts, M., & Niemivirta, M. (2005). Self-Regulated Learning: Finding a Balance between Learning Goals and Ego-Protective Goals. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 417-450). London: Elsevier Academic Press.

Bong, M., & Clark, R. E. (1999). Comparison Between Self-concept and Self-Efficacy in Academic Motivation Research. *Educational Psychologist, 34*(3), 139-153.

Borg, W. R., & Gall, M. D. (1989). *Educational Research: An Introduction* (5 ed.). New York: Longman.

Bourque, L. M. (1999). The Role of the National Assessment of Educational Progress (NAEP) in Setting, Reflecting, and Linking National Education Policy to States' Needs. In G. Cizek, J., (Ed.), *Handbook of Educational Policy* (pp. 213-249). California, U.S.A.: Academic Press.

Bowen, N. K., & Lee, J.-S. (2006). Parental Involvement, Cultural Capital, and the Achievement Gap Among Elementary School Children. *American Educational Research Journal, 43*(2), 193-218.

Brand, S., Felner, R., Shim, M., Seitsinger, A., & Dumas, T. (2003). Middel School Improvement and reform: Development and Validation of a School-Level Assessment of Climate, Cultural Pluralism, and School Safety. *Journal of Educational Psychology, 95*(3), 570-588.

Braymen, R. K. F., & Piersel, W. C. (1987). The Early Entrance Option: Academic and Social/Emotional Outcomes. *Psychology in Schools, 24*(2), 179-189.

Bridgman, P., & Davis, G. (2007). *The Australian Policy Handbook* (4 ed.). Crows Nest: Allen & Unwin.

Brock, C. H., Lapp, D., Flood, J., Fisher, D., & Han, K. T. (2007). Does Homework Matter? An Investigation of Teacher Perceptions About Homework Practices for Children From Nondominant Backgrounds. *Urban Education, 42*(4), 349-372.

Brookover, W., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). *School Social Systems and Student Achievement: Schools Can Make a Difference*. New York: Praeger.

Buhs, E. S., Ladd, G. W., & Herald, S. L. (2006). Peer Exclusion and Victimization: Processes That Mediate the Relation Between Peer Group Rejection and Children's Classroom Engagement and Achievement? *Journal of Educational Psychology, 98*(1), 1-13.

Byrne, B. M. (1998). *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming*. New Jersey: Lawrence Erlbaum Associates.

Caldas, S. J., & Bankston(III), C. (1997). Effect of School Population: Socioeconomic Status on Individual Academic Achievement. *The Journal of Educational Research, 90*(5), 269-277.

Caldwell, B. J. (2005). School-based Management, Education Policy Series Available from http://smec.curtin.edu.au/local/documents/Edpol3.pdf

Cameron, M. B., & Wilson, B. J. (1990). The Effects of Chronological Age, Gender, and Delay of Entry on Academic Achievement and Retention: Implications fro Academic Redshirting. *Psychology in Schools, 27*(3), 261-263.

Caprara, G. V., Barbaranelli, C., Borgogni, L., & Steca, P. (2003). Efficacy Beliefs as Determinants of Teachers' Job Satisfaction. *Journal of Educational Psychology, 95*(4), 821-832.

CAPSD, & BBE. (2007). *Syllabus for 9 & 10: Bhutan Certificate for Secondary Education*. Retrieved from http://www.education.gov.bt/Secretariat/syllabus.htm.

Carver, C. S., & Scheier, M. F. (2005). On the Structure of Behavioural Self-Regulation. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of Self-Regulation* (2nd ed., pp. 41-84). London: Elsevier Academic Press.

Casey, R. H., Gentile, P., & Biger, S. W. (1997). Teaching Appraisal in Higher Education. *Higher Education, 34*(4), 459-482.

Cizek, G. J. (2001). More Unintended Consequences of High-Stakes Testing. *Educational Measurement: Issues Practices, 20*(4), 19-27.

Cizek, G. J. (2005). High-Stakes Testing: Contexts, Characteristics, Critiques, and Consequences. In R. Phelps, P., (Ed.), *Defending Standardized Testing* (pp. 23-54). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Clark, M. A., Lee, S. M., Goodman, W., & Yacco, S. (2008). Examining Male Underachievement in Public Education: Action Research at a District Level. *NASSP Bulletin, 92*(2), 111-132.

Clark, M. A., Thompson, P., & Vialle, W. (2008). Examining the Gender Gap in Education Outcomes in Public Education: Involving Pre-Service School Counsellors and Teachers in Cross-Cultural and Interdisciplinary Research. *International Journal for Advanced Counselling, 30*(1), 55-66.

Clemant, M., & Vandenberghe, R. (2000). Teachers' professional development: a solitary or collegial (ad) venture? *Teaching and Teacher Education, 16*(1), 81-101.

Cochran, W. G. (1977). *Sampling Techniques* (3 ed.). New York: John Willey & Sons.

Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological Testing and Assessment: An Introduction to Test and Measurement* (4 ed.). Mountain View CA: Mayfield Publishing Company.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weindfeld, F. D., et al. (1966). *Equality of Educational Opportunity*. Washington, D.C., U.S.A: National Center for Educational Statistics.

Cook, T. D. (1985). Postpositivist Critical Multiplism. In R. L. Shortland & M. M. Mark (Eds.), *Social Science and Social Policy* (pp. 21-62). London: SAGE Publications, Inc.

Coombs, P. H. (1982). Critical World Educational Issues of the Next Two Decades. *International Review of Education, 28*(2), 143-157.

Cooper, H. (1989). Synthesis of Research on Homework. *Educational Leadership, 47*(3), 85-91.

Cooper, H., Lindsay, J. J., & Nye, B. (1998). Relationships Among Attitudes About Homework, Amount of Homework Assigned and Completed, and Student Achievement. *Journal of Educational Psychology, 90*(1), 70-83.

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does Homework Improve Academic Achievement? A Synthesis of Research, 1987-2003. *Review of Educational Research, 76*(1), pp. 1-62.

Cooper, H., & Valentine, J. C. (2001). Using Research to Answer Practical Questions About Homework. *Educational Psychologist, 36*(3), 143-153.

Cothran, D. J., Kulinna, P. H., & Garrahy, D. A. (2003). "This is kind of giving a secret away...": students' perspectives on effective class management. *Teaching and Teacher Education, 19*(4), 435-444.

Covington, M. V., & Omelich, C. L. (1984). Task-Oriented Versus Competitive Learnng Structures: Motivational and Performance Consequences. *Journal of Educational Psychology, 76*(6), 1038-1050.

Cox, P. J. (2005). *Participation and Performance in Mathematics and Science: Gender Issues Revisited*. Unpublished Research, La Trobe University, Victoria, Australia.

Creemers, B. P. M. (1994). *The Effective Classroom*. London: Cassell.

Creemers, B. P. M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 347-366.

Creemers, B. P. M., & Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice and Theory in Contemporary Schools*. London: Routledge Taylor & Francis Group.

Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (2 ed.). Thousand Oaks: Sage Publications, Inc.

Creswell, J. W. (2005). *Educational Research: Planning, Conducting, and Evaluating Quallitative and Quantitative Research* (2 ed.). New Jersey: Pearson.

Crocker, L., & Algina, J. (2008). *Introduction to Classical and Modern Test Theory*. New York: Rinehart and Winston.

Cropley, A. J. (1981). Lifelong learning: a rationale for teacher training. *Journal of Education in Teaching, 7*(1), 57-69.

Crosnoe, R., Riegle-Crumb, C., Frank, K., Field, S., & Muller, C. (2008). Peer Group of Girls' and Boys' Academic Experiences. *Child Development, 79*(1), 139-155.

Crum, K. S., & Sherman, W. H. (2008). Facilitating high achievement: High school principals' reflections on their successful leadership practices. *Journal of Educational Administration, 46*(5), 562-580.

Cullen, J. B., Jacob, B. A., & Levitt, S. D. (2005). The impact of school choice on student outcomes: an analysis of the Chicago Public Schools. *Journal of Public Economics, 89*, 729-760.

Daro, P., Stancavage, F., Ortega, M., DesStefano, L., & Linn, R. (2007). Validity Study of the NAEP Mathematics Assessment: Grades 4 and 8

Day, C. (2005). Sustaining success in challenging contexts: leadership in English schools. *Journal of Educational Administration, 43*(6), 573-583.

Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. *The Journal of Human Resources, 42*(3), 528-554.

Department of Education. (1999). *Education Sector Strategy: Realising Vision 2020 Policy and Strategy*. Retrieved from http://www.education.gov.bt/.

Desimone, L. (1999). Linking Parent Involvement With Student Achievement: Do Race and Income Matter? *The Journal of Educational Research, 93*(1), 11.

Desimone, L., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of Professional Development on Teachers' Instruction: Results from a Three-year Longitudinal Study. *Educational Evaluation and Policy Analysis, 24*(2), 81-112.

Diamantopoulos, A., & Siguaw, J. A. (2000). *Introducing LISREL*. Thousand Oaks: SAGE Publications.

Dohn, N. B. (2007). Knowledge and Skills for PISA-Assessing the Assessment. *Journal of Philosophy of Education, 41*(1), 1-16.

Donaldson, L. (2001). *The Contingency Theory of Organizations*. Thousand Oaks: Sage Publications, Inc.

Downey, D. B., Hippel, P. T. v., & Broh, B. A. (2004). Are Schools the Great Equalizer? Cognitive Inequality during the Summer Months and the School Year. *American Sociological Review, 69*(5), 613-635.

Downing, S. M. (2006). Twelve Steps for Effective Test Development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 3-25). New Jersey: Lawrence Erlbaum, Inc.

Doyle, W. (1980). *Classroom Management*. West Lafayette, Indiana: Kappa delta Pi.

Drent, M., & Meelissen, M. (2008). Which factors obstruct or stimulate teacher educators to use ICT innovatively? *Computers & Education, 51*(1), 187-199.

Driessen, G., Smit, F., & Sleegers, P. (2005). Parental Involvement and Educational Achievement. *British Educational Research Journal, 31*(4), 509-532.

Dweck, C. S. (1986). Motivational Processes Affecting Learning. *American Psychologist, 41*(10), 1040-1048.

Eamon, M. K. (2005). Social-Demographic, School, Neighbourhood, and Parenting Influences on the Academic Achievement of Latino Young Adolescents. *Journal of Youth and Adolescence, 34*(2), 163-174.

Eccles, J. S., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Reviews of Psychology, 53*, 109-132.

Ediger, M. (1996). Cooperative learning versus competition: Which is better? *Journal of Instructional Psychology, 23*, 204-300.

Education Monitoring & Support Services Division. (2003). *Factors Contributing to Classroom Effectiveness: A Sudy Report-2001*. Thimphu: Ministry of Education, Bhutan.

Educational Initiatives. (2009). *Annual Status of Student Learning*. Ahamedabad: Educational Initiatives Pvt. Ltd.

Egelund, N. (2008). The value of international comparative studies of achievement-a Danish perspective. *Assessment in Education: Principles, Policy & Practice, 15*(3), 245-251.

Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do Teachers' Race, Gender, and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labour Relations Review, 48*(3), 547-561.

Eide, E. R., & Goldhaber, D. D. (2005). Grade Retention: What Are the Costs and Benefits? *Journal of Education Finance, 31*(2), 195-214.

Ellington, A. J. (2003). A Meta-Analysis of the Effects of Calculators on Students' Achievement and Attitude Levels in Precollege Mathematics Classes. *Journal for Research in Mathematics Education, 34*(5), 433-463.

Elliott, J., Hufton, N., & Illushin, L. (2000). International Comparisons-What really matters? In D. Taylor, Shorrocks, & E. Jenkins, W., (Eds.), *Learning from Others* (Vol. 8, pp. 79-114). The Netherlands: Kluwer Academic Publishers.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Emmer, E. T., Evertson, C. M., Sandford, J. P., Clements, B. S., & Worsham, M. E. (1984). *Classroom Management for Secondary Teachers* (2 ed.). New Jersey: Prentice-Hall, Inc.

Eng, J. (2003). Sample Size Estimation: How Many Individuals Should Be Studied? *Radiology, 227*(2), 309-313.

Epstein, J. L. (1995). School/Family/Community Partnerships. *Phi Delta Kappan, 76*(9), 701-712.

Epstein, J. L., & Sanders, M. G. (2006). Connecting Home, School, and Community. In M. T. Hallinan (Ed.), *Handbook of the Sociology of Education* (pp. 285-306). New York: Springer.

Epstein, J. L., & VanVoorhis, F. L. (2001). More Than Minutes: Teachers' Roles in Designing Homework. *Educational Psychologist, 36*(3), 181-193.

Ertl, H. (2006). Educational Standards and the changing discourse on education: the reception and consequences of the PISA study in Germany. *Oxford Review of Education, 32*(5), 619-634.

Eurydice. (2007). *School Autonomy in Europe: Policies and Measures*. Brussels: European Unit.

Evers, A. (2001). The Revised Dutch Rating System for Test Quality. *International Journal of Testing, 1*(2), 155-182.

Fan, X., & Chen, M. (2001). Parental Involvement and Students' Academic Achievement: A Meta-Analysis. *Educational Psychology Review, 13*(1), 1-22.

Fensham, P. J. (2002). *Pointing the way forward for school science.* Paper presented at the Research Conference 2002 . Australian Council for Educational Research, Australia.

Finn, J. D. (1993). *School engagement and students at risk.* Washington, DC: National Center for Educational Statistics.

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and Inattentative-Withdrawn Behaviour and Achievement among Fourth Graders. *The Elementary School Journal, 95*(5), 421-434.

Finn, J. D., & Rock, D. A. (1997). Academic Success Among Students at Risk for School Failure. *Journal of Applied Psychology, 82*(2), 221-234.

Finn, J. D., & Voelkl, K. E. (1993). School Characteristics Related to Student Engagement. *The Journal of Negro Education, 62*(3), 249-268.

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research, 18*(1), 39-50.

Foster, C. A., & Marquart, D. J. (1984). *Performance-Based Funding in Public Schools* (No. ED 257826). Sacramento, CA: Sequoia Institute.

Fowler, F. J. (2009). *Survey Research Methods* (4 ed.). Thousand Oaks: Sage Publications, Inc.

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. (1987). Syntheses of Educational Productivity Research. *International Journal of Educational Research, 11*, 145-252.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research, 74*(1), 59-109.

Fuchs, T., & Wobmann, L. (2007). What accounts for international difference in student performance? A re-examination using PISA data. *Empirical Economics, 32*(2-3), 433-464.

Fuller, B., Gesicki, K., Kang, E., & Wright, J. (2006). Is the No Child Left Behind Act Working? The Reliability of How States Track Achievement. Working Paper 06-1. 49. Retrieved from http://eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED492024&ERICExtSearch_SearchType_0=no&accno=ED492024

Furner, J. M., & Berman, B. T. (2003). Math Anxiety: Overcoming a Major Obstacle to the Improvement of Student Math Performance. *Childhood Education, 79*(3), 170-174.

Gaer, E. V. d., Pustjens, H., Damme, J. V., & Munter, A. D. (2006). The Gender Gap in Language Achievement: The Role of School-Related Attitudes of Class Groups. *Sex Roles, 55*(5-6), 397-408.

Gamoran, A. (1992). Synthesis of Research: Is Ability Grouping Equitable? *Educational Leadership, 50*(2), 11-17.

Ganzeboom, H. B. G., Graaf, P. M. D., Treiman, D. J., & Leeuw, J. D. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research, 21*(1), 1-56.

Ganzeboom, H. B. G., & Treiman, D. J. (1996). Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research, 25*(3), 201-239.

Geist, E. (2010). The Anti_Anxiety Curriculum: Combating Math Anxiety in the Classroom. *Journal of Instructional Psychology, 37*(1), 24-31.

Gibbons, S., Machin, S., & Silva, O. (2008). Choice, competition, and pupil achievement. *Journal of European Economic Association, 6*(4), 912-947.

Gibbons, S., & Telhaj, S. (2007). Are Schools Drifting Apart? Intake Stratification in English Secondary Schools. *Urban Studies, 44*, 1281-1305.

GNH Commission. (2008). *Draft Tenth Five Year Plan (2008-2013) Vol I: Main Document.* Retrieved from http://www.pc.gov.bt/fyp/Draft%2010thplan/Draft_Main_Document_Volume_I.pdf.

Goddard, R. D., & Goddard, Y. L. (2001). A multilevel analysis of the relationship between teacher and collective efficacy in urban schools. *Teaching and Teacher Education, 17*, 807-818.

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective Teacher Efficacy: Its Meaning, Measure, and Impact on Student Achievement. *American Educational Research Journal, 37*(2), 479-507.

Goldstein, H. (2004). International comparisons of student attainment:some issues arising from the PISA study. *Assessment in Education: Principles, Policy & Practice, 11*(3), 319-330.

Goldstein, H., Bonnet, G., & Rocher, T. (2007). Multilevel Structural Equation Models for the Analysis of comparative data on educational Performance. *Journal of Educational and Behavioural Statistics*, *32*(3), 252-286.

Goldstein, H., & Leckie, G. (2008). School league tables: what can they really tell us? *Significance, 5*(2), 67-69.

Goldstein, H., & Thomas, S. M. (2008). Reflection on the international comparative surveys debate. *Assessment in Education: Principles, Policy & Practice, 15*(3), 215-222.

Good, T. L. (1987). Two Decades of Research on Teacher Expectations: Findings and Future Directions. *Journal of Teacher Education, 38*(4), 32-47.

Goodman, D., & Hambleton, R. K. (2005). Some Misconceptions About Large-Scale Educational Assessments. In R. Phelps, P., (Ed.), *Defending Standardized Testing* (pp. 91-110). New Jersey: Lawrence Elbaum Associates, Inc., Publishers.

Graham, J., Tisher, R., Ainley, M., & Kennedy, G. (2008). Staying with the text: the contribution of gender, achievement orientations, and interest to students' performance on a literacy task. *Educational Psychology, 28*(7), 757-776.

Gratton, R. (2004). Teacher Appraisal: a lesson on confusion over purpose. *International Journal of Educational Management, 18*(5), 292-296.

Grauwe, A. D. (2005). Improving the Quality of Education through Schol-Based management: Learning from International Experiences. *Internal Review of Education*, *51*(4), 269-287.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis, 11*(3), 255-274.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The Effect of School Resources on Student Achievement. *Review of Educational Research, 66*(3), 361-396.

Greenwood, C. R. (1991). Longitudinal Analysis of Time, Engagement, and Achievement in At-Risk Versus Non-Risk Students. *Exceptional Children, 57*(6), 521-535.

Grissmer, D. (2002). *Improving NAEP for Research and Policymaking*. Washington, D.C: National Assessment Governing Board.

Grolnick, W. S., Benjet, C., Kurowski, C. O., & Apostoleris, N. H. (1997). Predictors of Parent Involvement in Children's Schooling. *Journal of Educational Psychology, 89*(3), 538-549.

Grolnick, W. S., & Slowiaczek, M. L. (1994). Parents' Involvement in Children's Schooling: A Multidimensional Conceptualization and Motivational Model. *Child Development, 65*(1), 237-252.

Gronlund, N. E., & Waugh, C. K. (2009). *Assessment of Student Achievement* (9 ed.). New Jersey: Upper Saddle River.

Guba, E. G. (1990). The Alternative Paradigm Dialog. In E. G. Guba (Ed.), *The Paradigm Dialog* (pp. 17-30). Newburry Park: SAGE Publications, Inc.

Guba, E. G., & Lincoln, Y. S. (1994). Competing Paradigms in Qualitative Research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 105-117). Thousand Oaks SAGE Publications, Inc.

Guevremont, A., Ross, N. P., & Brownell, M. (2007). Predictors and Consequences of Grade Retention. *Canadian Journal of School Psychology, 22*(1), 50-67.

Gurr, D., Drysdale, L., & Mulford, B. (2005). Successful principal leadership: Australian case studies. *Journal of Educational Administration, 43*(6), 539-551.

Hallinger, P. (1989). Developing Instructional Leadership Teams in Secondary Schools: a Framework. *NASSP Bulletin, 73*(517), 84-92.

Hallinger, P. (1990). Principal Instructional Management Rating Scale: Principal Form 2.0. Sarasota, FL: Leading Development Associates.

Hallinger, P. (1994). Introduction: Exploring the Impact of Principal Leadership. *School Effectiveness and School Improvement, 5*(3), 206-218.
Hallinger, P. (2003). Leading Educational Change: reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education, 33*(3), 329-352.
Hallinger, P. (2005). Instructional Leadership and the School Principal: A Passing Fancy that Refuses to Fade Away. *Leadership and Policy in Schools, 4*, 221-239.
Hallinger, P., & Heck, R. H. (1996). Reassessing the Principal's Role in School Effectiveness: A Review of Empirical Research, 1980-1995. *Educational Administration Quarterly, 32*(1), 5-44.
Hallinger, P., & Heck, R. H. (1998). Exploring the Principal's Contribution to School Effectiveness: 1980-1995. *School Effectiveness and School Improvement, 9*(2), 157-191.
Hamp-Lyons, L. (1997). Washback, Impact and Validity: ethical concerns. *Language Testing, 14*(3), 295-303.
Hanushek, E. A. (1979). Conceptual and Empirical Issues in the Estimation of Educational Production Functions. *The Journal of Human Resources, 14*(3), 351-388.
Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature, 24*(3), 1141-1177.
Hanushek, E. A. (1989). The Impact of Differential Expenditures on School Performance. *Educational Researcher, 18*, 17-26.
Hanushek, E. A. (1997). Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis, 19*(2), 141-164.
Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Disruption versus Tiebout improvement: the costs and benefits of switching schools. *Journal of Public Economics, 88*, 1721-1746.
Hanushek, E. A., & Wobmann, L. (2006). Does Educational Tracking Affect Performance and Inequality? Differences-In-Diferences Evidence Across Countries. *The Economic Journal, 116*, 63-76.
Hargreaves, A. (2001). The emotional geographies of teachers' relations with colleagues. *International Journal of Educational Research, 35*(5), 503-527.
Harlen, W. (2001). Scientific Literacy-Conceptions and Assessment. In H. Behrendt & et al. (Eds.), *Research in Science Education-Past, Present, and Future* (pp. 49-60). The Netherlands: Kluwer Academic Publishers.
Harris, A. (1998). Effective Teaching: a review of the literature. *School Leadership and Management, 18*(2), 169-183.
Hart, P. M., Wearing, A. J., Conn, M., Carter, N. L., & Dingle, R. K. (2000). Development of the School Organizational Health Questionnaire: A measure for assessing teacher morale and school organistional climate. *British journal of educational Psychology, 70*(2), 211-228.
Hattie, J. (1985). Methology Review: Assessing Unidimensionality of Tests and Items. *Applied Psychological Measurement, 9*(2), 139-164.
Hattie, J. (2009). *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*. New York: Routledge Taylor & Francis Group.
Hatzinikita, V., Dimopoulos, K., & Christidou, V. (2008). PISA test items and school textbooks related to science: A textual comparison. *Science Education, 92*(4), 567-590.
Hauck, A. L., & Finch(Jr), A. J. (1993). The Effect of Relative Age on Achievement in Middle School. *Psychology in Schools, 30*(1), 74-79.
Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes. *Educational Researcher, 23*(3), 5-14.
Hellum-Alexander, A. (2010). *Effective Teaching Strategies for Alleviating Math Anxiety and Increasing Self-Efficacy in Secondary Students.* Unpublished Master's Thesis, The Evergreen College, Washington.
Hembree, R. (1990). The Nature, Effects, and Relief of Mathematics Anxiety. *Journal for Research in Mathematics Education, 21*(1), 33-46.

Heuvelmans, T. (2002). TiaPlus (Version 2.1). Arnhem, NL: Cito, Measurement and Research Department.

Hidi, S., & Harackiewickz, J. M. (2000). Motivating the Academically Unmotivated: A Critical Issue for the 21st Century. *Review of Educational Research*, *70*(2), 151-179.

Ho, H.-Z., Senturk, D., Lam, A. G., Zimmer, J. M., Hong, S., & Okamoto, Y. (2000). The Affective and Cognitive Dimensins of Mth Anxiety: A Cross-National Study. *Journal for Research in Mathematics Education, 31*(2), 362-379.

Hoog, J., Johansson, O., & Olofsson, A. (2005). Successful principalship: the Swedish case. *Journal of Educational Administration, 43*(6), 595-606.

Hoogland, J. J., & Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis. *Sociological Methods & Research, 26*(3), 329-367.

Hoover-Dempsey, K. V., Battiaoto, A. C., Walker, J. M. T., Reed, R. P., DeJong, J. M., & Jones, K. P. (2001). Parental Involvement in Homework. *Educational Psychologist, 36*(3), 195-209.

Hoover-Dempsey, K. V., & Sandler, H. M. (1995). Parental Involvement in Children's Education: Why Does It Make a Difference? *Teachers College Record, 97*(2), 311-331.

Hoover-Dempsey, K. V., & Sandler, H. M. (1997). Why Do Parents Become Involved in Their Children's Education? *Review of Educational Research, 67*(1), 3-42.

Houtte, M. V. (2004a). Gender context of the school and study culture, or how the presence of girls affects the achievement of boys. *Educational Studies, 30*(4), 404-423.

Houtte, M. V. (2004b). Why boys achieve less at school than girls: the difference between boys' and girls' academic culture. *Educational Studies, 30*(2), 159-173.

Howe, K. R. (1992). Getting over the Quantitative-Qualitative Debate. *American Journal of Education, 100*(2), 236-256.

Howie, J. S. (2000). TIMSS in South Africa *The Value of International Comparative Studies for a Developing Country*. In D. Taylor, Shorrocks, & E. Jenkins, W., (Eds.), *Learning from Others* (pp. 279-301). The Netherlands: Kluwer Academic Publishers.

Hoy, W. K., & Hannum, J. W. (1997). Middle School Climate: An Empirical Assessment of Organizational Health and Student Achievement. *Educational Administration Quarterly, 33*(3), 290-311.

Hughes, J. N., Luo, W., Kwok, O.-M., & Loyd, L. K. (2008). Teacher-Student Support, Effortful Engagement, and Achievement: A 3-Year longitudinal Study. *Journal of Educational Psychology, 100*(1), 1-14.

Husen, T., & Postlethwaite, T. N. (1996). A Brief History of the International Association for the Evaluation of Educational Achievement (TEA). *Assessment in Education: Principles, Policy & Practice, 3*(2), 129-141.

Hussein, M. G. (1992). What does Kuwait want to learn from the Third international Mathematics and Science Study (TIMSS) ? *Prospects, 22*(4), 463-468.

iDiscoveri Education, & Royal Education Council. (2009). *The Quality of School Education in Bhutan : Reality and Opportunities*. Retrieved from http://www.rec.org.bt/formdownload.htm.

IEA. (1998). *IEA Trends in Internatinal Mathematics and Science Study 1999 Mathematics Teacher Questionnaire*. Retrieved from http://timss.bc.edu/timss1999i/pdf/BM2_TeacherM.pdf.

IEA. (2000). *TIMSS 1999 Technical Report*. Boston, U.S.A.: International Study Center, Lynch School of Education, Boston College,.

IEA. (2004). *TIMSS 2003 Technical Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eight Grades*. Boston, U.S.A: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

IEA. (2008). *TIMSS 2007 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Centre, Lynch School of Education, Boston College.

Ilomaki, L., & Rantanen, P. (2007). Intensive use of ICT in school: Developing differences in students' ICT expertise. *Computers & Education, 48*, 119-136.

Ingvarson, L., & Rowe, K. (2007). *Conceptualising and Evaluating Teacher Quality: Substantive and Methodological Issues*. Paper presented at the Economics of Teacher Quality.

International Labour Office. (1990). *ISCO-88: International Standard Classification of Occupations*.

Izzo, C. V., Weissberg, R. P., Kasprow, W. J., & Fendrich, M. (1999). A Longitudinal Assessment of Teacher Perceptions of Parent Involvement in Children's Education and School Performance. *American Journal of Community Psychology, 27*(6), 1999.

Jacobson, S. L., Johnson, L., Ylimaki, R., & Giles, C. (2005). Successful leadership in challenging US schools: enabling principles, enabling schools. *Journal of Educational Administration, 43*(6), 607-618.

Jencks, C., Smith, M. S., Ackland, H., Bane, M. J., Cohen, D., Grintlis, H., et al. (1972). *A reassessment of the effect of family and schooling in America*. New York: Basic Books.

Jenkins, E. W. (2000). Making Use of International Comparisons of Student Achievement in Science and Mathematics. In D. Taylor, Shorrocks, & E. Jenkins, W. (Eds.), *Learning from Others* (pp. 137-157). The Netherlands: Kluwer Academic Publishers.

Jeynes, W. H. (2003). A Meta-Analysis: The Effects of Parental Involvement on Minority Children's Academic Achievement. *Education and Urban Society, 35*(2), 202-218.

Jeynes, W. H. (2005). A Meta-Analysis of the Relation of Parental Involvement to Urban Elementary School Student Academic Achievement. *Urban Education, 40*(3), 237-269.

Jeynes, W. H. (2007). The Relationship Between Parental Involvement and Urban Secondary School Student Academic Achievement: A Meta-Analysis. *Urban Education, 42*(1), 82-110.

Jimerson, S. R. (2001). Meta-analysis of Grade Retention Research: Implications for Practice in the 21st Century. *School Psychology Review, 30*(3), 420-437.

Jimerson, S. R., Anderson, G. E., & Whipple, A. D. (2002). Winning the Battle and Losing the War: Examining the Relation between Grade Retention and Dropping Out of High School. *Psychology in Schools, 39*(4), 441-457.

Jimerson, S. R., Pletcher, S. M. W., Graydon, K., Schnurr, B. L., Nickerson, A. B., & Kundert, D. K. (2006). Beyond Grade Retention and Social Promotion: Promoting the Social and Academic Competence of Students. *Psychology in Schools, 43*(1), 85-97.

Johnson, E. G. (1992). The Design of the National Assessment of Educational Progress. *Journal of Educational Measurement, 29*(2), 95-110.

Johnson, E. G., & Owen, E. (1998). *Linking The National Assessment of Educational Progress (NAEP) and The Third International Mathematics and Science Study (TIMSS): A Technical Report* (No. NCES 98-499). Washington, DC: National Center for Educational Statistics.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed-Methods Research: A Research Paradigm Whose TIme Has Come. *Educational Researcher, 33*(7), 14-26.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research, 1*(2), 112-133.

Joreskog, K., & Sorbom, D. (2002). *PRELIS 2: User's Reference Guide* (3 ed.). Lincolnwood: Scientific Software International, Inc.

Joreskog, K., & Sorbom, D. (2003). LISREL (Version 8.54). Lincolnwood: Scientific Software International, Inc.

Kafer, K. (2001). A Guide to the NAEP Achievement Test. Retrieved from http://www.heritage.org/Research/Education/BG1419.cfm

Kaplan, R. M., & Saccuzzo, D. P. (2001). *Psychological Testing: Principles , Applications, and Issues* (5 ed.). Belmont, CA: Wadsworht/ Thomson Learning.

Kerlinger, F. N. (1989). *Foundations of Behavioral Research* (3 ed.). New York: Holt, Rinehart & Winston.

Kincheloe, J. L., & McLaren, P. (2003). Rethinking of Critical Theory and Qualitative Research. In N. K. Denzen & Y. S. Lincoln (Eds.), *The Landscape of Qualitative Research: Theories and Issues* (2 ed., pp. 433-488). Thousand Oaks: SAGE Publications, Inc.

Kish, L. (1995). *Survey Sampling* (Classics ed.). New York: JOHN WILEY & SONS, INC.

Kitchen, A. (2000). The need for caution when using the results of international mathematics testing to inform policy decisions in education in the member countries. In D. Taylor, Shorrocks, &

E. Jenkin, W. (Eds.), *Learning from Others* (pp. 219-231). The Netherlands: Kluwer Academic Publishers.
Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2 ed.). New York: The Guilford Press.
Knapper, C. K., & Cropley, A. J. (2000). *Lifelong Learning in Higher Education* (3 ed.). London: Kogan Page Ltd.
Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking* (2 ed.). New York: Springer.
Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2008). A Multilevel Study of Predictors of Student Perceptions of School Climate: The Effect of Classroom-Level Factors. *Journal of Educational Psychology, 100*(1), 96-104.
Krathwohl, D. R. (2009). *Methods of Educational and Social Science Research* (3 ed.). Illinois: Waveland Press Inc.
Krueger, R. A., & Casey, M. A. (2009). *Focus Groups: A Practical Guide for Applied Research* (4 ed.). Thousand Oaks: SAGE Publications, Inc.
Kundert, D. K., May, D. C., & Brent, D. (1995). A Comparison of Students Who Delay Kindergarten Entry and Those Who are Retained in Grades K-5. *Psychology in Schools, 32*(3), 202-209.
Kyriakides, L. (2008). Testing the validity of the comprehensive model of educational effectiveness: a step towards the development of a dynamic model of effectiveness. *School Effectiveness and School Improvement, 19*(4), 429-446.
Lange, J. d. (2002). *The meaning of PISA for teachers of mathematics.* Paper presented at the ACER Research Conference 2002 Providing World-Class School Education: What does the research tell us?, Australian Council for Educational Research, Australia.
Lavy, V. (2010). Effects of Free Choice Among Public Schools. *Review of Economic Studies, 77*, 1164-1191.
Learly-Kelly, S. W. O., & Vokurka, R. J. (1998). The Empirical Assessment of Construct Validity. *Journal of Operations Management 16*, 387-405.
Lee, J., Grigg, W. S., & Dion, G. S. (2007). *The Nation's Report Card: Mathematics 2007*. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007494.
Leimu, K. (1992). Interest and modes in research utilization: the Finnish IEA experience. *Prospects, 22*(4), 426-433.
Leithwood, K., & Jantzi, D. (1990). Transformational Leadership: How Principals Can Help Reform School Cultures. *School Effectiveness and School Improvement, 1*(4), 249-280.
Leithwood, K., & Jantzi, D. (2005). A Review of Transformational School Leadership Research 1996-2005. *Leadership and Policy in Schools, 4*(3), 177-199.
Lenth, R. V. (2001). Some Practical Guidelines for Effective Sample Size Determination. *The American Statistician, 55*(3), 187-193.
Lie, S. (2005). How can Large International Comparative Studies Contribute to the Quality of Science Education? In K. Boersma, M. Goedhart, O. Jong, De, & H. Eijkelhof (Eds.), *Research and the Quality of Science Education* (pp. 27-40). Dordrecht, The Netherlands: Springer.
Lim, C. P. (2002). A theoretical framework for the study of ICT in schools: a proposal. *British Journal of Educational Technology, 33*(4), 411-421.
Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Newbury Park: SAGE Publications, Inc.
Lindquist, M. M. (2001). NAEP, TIMSS, and PSSM: Entangled Influences. *School Science and Mathematics, 101*(6).
Lindquist, M. M. (Ed.). (1989). *Fourth Mathematics Assessment of Educational Progress of the National Asessment of Educational Progress*. Virginia: The National Council of teachers of Mathematics, Inc.
Linn, R. L. (2002). Validation of the Uses and Interpretations of Results of State Assessment and Accountability Systems. In G. Tindal & T. M. Haladyna (Eds.), *Large-Scale Assessment Programs for All Students: Validity, Technical Adequacy, and Implementation* (pp. 27-48). London: Lawrence Erlbaum Associates, Publishers.

Lisle, J. D., Smith, P., & Jules, V. (2005). Which males or females are more at risk and on what? An analysis of gender differentials within the primary school system of Trinidad and Tobago. *Educational Studies, 31*(4), 393-418.

Little, S. G., & Akin-Little, A. (2008). Psychology's Contributions to Classroom Management. *Psychology in the Schools, 45*(3), 227-234.

Lokan, J. (2000). Messages for Mathematics Education from TIMSS in Australia. In K. Tobin (Ed.), *Learning from Others* (pp. 259-277). The Netherlands: Kluwer Academic Publishers.

Lopez, O. S. (2007). Classroom Diversification: A Strategic View of Educational Productivity. *Review of Educational Research, 77*(1), 28-80.

Lumsden, L. (1998). Teacher Morale. *ERIC Digest, EDO-EA-98-4*(120). Retrieved from httpi/erie.uorgon.edu.publications/ digests/ digest120.html

Lynch, K., & Baker, J. (2005). Equality in Education: An Equality of Condition Perspective. *Theory and Research in Education, 3*(2), 131-164.

Ma, X. (1999). A Meta-Analysis of the Relationship Between Anxiety Towards Mathematics and Achievement in Mathematics. *Journal for Research in Mathematics Education, 30*(5), 520-540.

Mackenzie, N. (2007). Teacher morale: More complex than we think? *The Australian Educational Researcher, 34*(1), 89-104.

Macneil, A. J., Prater, D. L., & Bush, S. (2007). The effects of school culture and climate on student achievement. Retrieved from http://dx.doi.org/10.1080/13603120701576241. doi:10.1080/13603120701576241

Maeyer, S. D., Rymenans, R., Petegem, P. V., Berg, H. v. d., & Rijlaarsdam, G. (2007). Educational Leadership and Pupil Achievement: The choice of a valid conceptual model to test effects in school effectiveness research. *School Effectiveness and School Improvement, 18*(2), 125-145.

Malacova, E. (2007). Effect of single-sex education on progress in GCSE. *Oxford Review of Education, 33*(2), 233-259.

Marks, H. M., & Printy, S. M. (2003). Principal Leadership and school Performance: An Integration of Transformational and Instructional Leadership. *Educational Administration Quarterly, 39*(3), 370-397.

Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Science Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eight Grades*. Boston, U.S.A.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Marzano, R. J. (1998). *A Theory-Based Meta-Analysis of Research on Instruction*. Aurora, CO: Mid-continent Regional Educational Laboratory.

Marzano, R. J. (2001). *A New Era of School Reform: Going Where the Research Takes Us*. Washington, DC: Mid-continent Research for Education and Learning, 2550 S., Parker Road, Aurora.

Marzano, R. J. (2003a). *Classroom Management that Works: Research-Based Strategies for Every Teacher*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2003b). *What Works in Schools: Translating Research into Action*. USA: Association for Supervision and Curriculum Development.

Marzano, R. J., Gaddy, B. B., & Dean, C. (2000). *What Works In Classroom Instruction*. Aurora: Mid-continent Research for Education and Learning.

Maslowski, R., Scheerens, J., & Luyten, H. (2007). The Effect of School Autonomy and School Internal Decentralization on Students' Reading Literacy. *School Effectiveness and School Improvement, 18*(3), 303-334.

Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika, 47*(2), 149-174.

Masters, G. N., & Wright, B. D. (1997). The Partial Credit Model. In W. J. v. d. Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 101-121). New York: Springer.

Maykut, P., & Morehouse, R. (1994). *Beginning Qualitative Research: A Philosophic and Practical Guide*. London: The Palmer Press.
McGaw, B. (2008a). Further reflections. *Assessment in Education: Principles, Policy & Practice, 15*(3), 279-282.
McGaw, B. (2008b). The role of the OECD in international comparative studies of achievement. *Assessment in Education: Principles, Policy & Practice, 15*(3), 223-243.
Michael, O. L. (2001). The Effects of Age-based and Grade-based Sampling on the Relative Standing of Countries in International Comparative Studies of Student Achievement. *British Educational Research Journal, 27*(2), 187-200.
Ministry of Education. (2006). *Quality of Education (Standards)*. Retrieved from http://www.education.gov.bt/CCM%20presentation%20on%20Quality%20of%20education.pdf.
Ministry of Education. (2010). *Teacher Performance Appraisal: Technical Requirements Manual*. Retrieved from http://www.edu.gov.on.ca/eng/teacher/pdfs/TPA_Manual_English_september2010l.pdf.
Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating Population Characteristics From Sparse Matrix Samples of Item Responses. *Journal of Educational Measurement, 29*(2), 133-161.
Moller, J., Eggen, A., Fuglestad, O. L., Langfeldt, G., Presthus, A.-M., Skrovset, S., et al. (2005). Successful school leadership: the Norwegian case. *Journal of Educational Administration, 43*(6), 584-594.
Monk, D. H. (1992). Educational Evaluation and Policy Analysis. *Educational Evaluation and Policy Analysis, 14*(4), 307-332.
Moos, L., Krejsler, J., Kofod, K. K., & Jensen, B. B. (2005). Successful school principalship in Danish schools. *Journal of Educational Administration, 43*(6), 563-572.
Morgan, D. L. (1988). *Focus Groups as Qualitative Research*. Newbury Park: SAGE Publications Inc.
Morgan, D. L., & Scannell, A. U. (1998). *Planning Focus Groups: Focus Group Kit 2*. Thousand Oaks: SAGE Publications.
Mortimore, P. (1993). School Effectiveness and the Management of Effective Learning and Teaching. *School Effectiveness and School Improvement, 4*(4), 290-310.
Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. J. (1988). *School Matters: The JuniorYears*. Wells: Open Books.
Muhlenbruck, L., Cooper, H., & Nye, B. (2000). Homework and achievement: Explaining the different strengths of relation at the elementary and secondary school levels. *Social Psychology of Education, 3*, 295-317.
Mullis, I. V. S. (2002). *Background Questions in TIMSS and PIRLS: An Overview*. Washington, D.C: National Assessment Governing Board.
Mullis, I. V. S., Kennedy, A. M., Martin, M. O., & Foy, P. (2007). PIRLS 2006 International Report Available from http://pirls.bc.edu/pirls2006/intl_rpt.html
Mullis, I. V. S., Kennedy, A. M., Martin, M. O., & Sainbury, M. (2006). PIRLS 2006 Assessment Framework and Specifications Available from http://pirls.bc.edu/pirls2006/framework.html
Mullis, I. V. S., Martin, M. O., Foy, P., Olson, J. F., Preuschoof, C., Erberber, E., et al. (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A., & Erberber, E. (2005). TIMSS 2007 Assessment Frameworks Available from http://pirls.bc.edu/TIMSS2007/frameworks.html
Mullis, I. V. S., Martin, M. O., Ruddock, G. J., Sullivan, C. Y. O., & Preuschoff, C. (2009). *TIMSS 2011: Assessment Frameworks*. Chestnut Hill: TIMSS and PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Smith, T. A., Garden, R. A., Gregory, R. A., Gonzalez, E. J., et al. (2003). TIMSS Assessment Frameworks and Specifications 2003 Available from http://pirls.bc.edu/timss2003i/frameworksD.html

Mullis, I. V. S., Martin, M. O., Smith, T. A., Garden, R. A., Gregory, R. D., Gonzalez, E. J., et al. (2000a). *TIMSS 1999 International Mathematics Report*. Boston, U.S.A.: International Study Center, Lunch School of Education, boston College.

Mullis, I. V. S., Martin, M. O., Smith, T. A., Garden, R. A., Gregory, R. D., Gonzalez, E. J., et al. (2000b). *TIMSS 1999 International Science Report*. Boston, U.S.A: International Study Center, Lynch School of Education, Boston College.

Mullis, I. V. S., Matin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Mathematics Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eight Grades*. Boston, U.S.A.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

NAGB. (1996). Report to the National Assessment Governing Board. Retrieved 20-06-2008, from http://www.nagb.org/pubs/appj.html

NAGB. (2000). Student Performance Standards on the National Assessment of Educational Progress: Affirmation and Improvements (2000). In M. Bourque, Lyn, & S. Byrd (Eds.) Available from http://nagb.org/

NAGB. (2001). *The NAEP 1998 Technical Report*. Washington, D.C: National Center for Education Statistics.

NAGB. (2003). Background Information Framework for the National Assessment of Educational Progress. Retrieved from http://www.nagb.org/

NAGB. (2004). *NAEP 2009 Science Framework Development: Issues and Recommendations*. Retrieved from http://www.nagb.org/release/iss_paper11_22_04.doc.

NAGB. (2005). *Science Framework for the 2005 National Assessment of Educational Progress*. Retrieved from http://nces.ed.gov/transfer.asp?location=nagb.org/pubs/s_framework_05/761907-ScienceFramework.pdf.

NAGB. (2006a). *Mathematics Framework for the 2007 National Assessment of Educational Progress*. Retrieved from http://nces.ed.gov/transfer.asp?location=www.nagb.org/frameworks/math_07.pdf.

NAGB. (2006b). The Nation's Report Card: Science 2005 Available from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006466

NAGB. (2007a). *Mathematics Framework for 2009 National Assessment of Educational Progress*. Retrieved from http://nagb.org/.

NAGB. (2007b). *Science Framework for the 2009 National Assessment of Educational Progress*. Retrieved from http://nagb.org/.

Nardi, E. (2008). Cultural biases: a non-Anglophone perspective. *Assessment in Education: Principles, Policy & Practice, 15*(3), 259-266.

National Center for Education Statistics. (2011). *The Nation's Report Card: Science 2009* (No. NCES 2011-451). Washington, D.C: Institute of Education Sciences, U.S. Department of Education.

NCTM. (2000). *Principles and Standards for School Mathematics*. Reston, VA: NCTM.

Newstead, K. (1998). Aspects of Children's mathematics Anxiety. *Educational Studies in Mathematics, 36*(1), 53-71.

Norwood, K. S. (1994). The Effect of Instructional Approach on Mathematics Anxiety and Achievement. *School Science and Mathematics, 94*(5), 248-254.

Nunnally, J. C. (1978). *Psychometric Theory* (2 ed.). New York: McGraw-Hill Book Company.

OECD. (1999a). *Classifying Educational Programmes: Manual for ISCED-97 Implementation in OECD Countries*. Retrieved from http://www.oecd.org/dataoecd/41/42/1841854.pdf.

OECD. (1999b). Measuring Student Knowledge and Skills: A New Framework for Assessmentpp. 85). Available from

http://lysander.sourceoecd.org/vl=6357152/cl=16/nw=1/rpsv/ij/oecdthemes/99980029/v1999n1/s1/p1
OECD. (2000). *Measuring Student Knowledge and Skills:The PISA 2000 Assessment of Reading, Mathematical and Scientific Literacy*. Paris: OECD.
OECD. (2001). Teachers for Tomorrow's Schools: Analysis of the World Education Indicators Available from http://lysander.sourceoecd.org/vl=1112734/cl=35/nw=1/rpsv/ij/oecdthemes/99980029/v2001n12/s1/p1l
OECD. (2002). *PISA 2000 Technical Report*. Paris: OECD.
OECD. (2004a). *Learning for Tomorrow's World: First Result from PISA 2003*. Paris: OECD.
OECD. (2004b). *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills* OECD, Paris.
OECD. (2004c). *Policy Brief: Lifelong Learning*. Retrieved from http://www.oecd.org/dataoecd/17/11/29478789.pdf.
OECD. (2004d). *What Makes School Systems Perform? Seeing School Systems through the Prism of PISA*. Paris: OECD.
OECD. (2005a). *Are Students Ready for a Technology-Rich World? What PISA Studies Tell Us*. Paris: OECD.
OECD. (2005b). Education Policy Analysis. Retrieved 20 August, 2008, from http://lysander.sourceoecd.org/vl=7441565/cl=11/nw=1/rpsv/ij/oecdthemes/99980029/v2005n9/s1/p1l
OECD. (2005c). *PISA 2003 Data Analysis Manual: SPSS* Paris: OECD.
OECD. (2005d). *PISA 2003 Technical Report*. Paris: OECD.
OECD. (2006). Assessing Scientific, Reading and Mathematical Literacy: A framework for PISA 2006 Available from http://www.oecd.org/pages/0,3417,en_32252351_32236191_1_1_1_1_1,00.html
OECD. (2007a). *PISA 2006: Science Competencies for Tomorrow's World: Volume 1 Analysis*. Paris: OECD.
OECD. (2007b). *PISA 2006: Volume 2: Data* OECD,Paris.
OECD. (2009a). *Creating Efective Teaching and Learning Environments: First Results From TALIS*. Paris: OECD.
OECD. (2009b). *PISA 2006 Technical Report*. Paris: OECD.
OECD. (2009c). *Take the Test: Sample Questions from OECD's PISA Assessments*. Paris: OECD publishing.
OECD. (2010). *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science* (Vol. 1). Paris: OECD Publishing.
Onwuegbuzie, A. J. (2002). Why can't we all get along? Towards a framework for unifying research paradigms. *Education, 122*(3), 518-530.
Orpwood, G. (2000). Diversity of Purpose in International Assessments *Issues arising from the TIMSS Tests of Mathematics and Science literacy*. In D. Taylor, Shorrocks, & E. Janovy, W. (Eds.), *Learning from Others* (pp. 49-62). The Netherlands: Kluwer Academic Publishers.
Pajares, F. (1996). Self-Efficacy Beliefs in Academic Settings. *Review of Educational Research, 66*(4), 543-578.
Parsons, W. (1995). *Public Policy: An Introduction to the Theory and Practice of Policy Analysis*. Cheltenham, U.K: Edward Elgar Publishing, Inc.
Pelgrum, W. J. (2001). Obstacles to the integration of ICT in education: results from a worldwide educational assessment. *Computers & Education, 37*(2), 163-178.
Phelps, R. P. (2005). The Rich, Robust Research Literature on Testing's Achievement Benefits. In R. Phelps, P., (Ed.), *Defending Standardized Testing* (pp. 55-90). New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.
Phillips, D. C. (1983). After the Wake: Postpositivistic Educational Thought. *Educational Researcher, 12*(5), 4-12.

Phillips, D. C. (1994). Postpositivistic Science. In E. G. Guba (Ed.), *The Paradigm Dialog*. Newbury Park: SAGE Publications, Inc.

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and Educational Research*. Maryland: Rowman & Littlefield Publishers, Inc.

Phillips, G. W. (2007). *Expressing International Educational Achievement in Terms of U.S. Performance Standards: Linking NAEP Achievement Levels to TIMSS*. Washington, D.C: American Institute for Research.

Pintrich, P. R., & De Groot, E., V. (1990). Motivational and Self-Regulated Learning Components of Classroom Academic Performance. *Journal of Educational Psychology, 82*(1), 33-40.

Pitner, N. J. (1988). The Study of Administrator Effects and Effectiveness. In N. J. Boyan (Ed.), *Handbook of Research on Educational Administration* (pp. 99-122). New York: Longman Inc.

Planning Commission. (1999). *Bhutan 2020: A Vision for Peace, Prosperity and Happiness Part II*. Retrieved from http://www.pc.gov.bt/publications/pub/Bhutan2020_2.pdf.

Policy and Planning Division. (2007). *General Statistics 2007*. Retrieved from http://www.education.gov.bt/.

Policy and Planning Division. (2008). *General Statistics 2008*. Retrieved from http://www.education.gov.bt/.

Policy and Planning Division. (2009). *Annual Education Statistics, 2009*. Retrieved from http://www.education.gov.bt/.

Policy and Planning Division. (2010). *Annual Education Statistics 2010*. Retrieved from http://www.education.gov.bt/Publications/AES%20Final.pdf.

Popkewitz, T. S. (1990). Whose Future? Whose Past? Notes on Critical Theory and Methogology. In E. G. Guba (Ed.), *The Paradigm Dialog* (pp. 46-66). Newbury Park: SAGE publications, Inc.

Prais, S. J. (2003). Cautions on OECD'S Recent Educational Survey (PISA) . *Oxford Review of Education*, *29*(2), 139-163.

Preston, R. (2008, 28-29 January). *Mathematics Anxiety: Research and Implications for Middle School Students and Teachers*. Paper presented at the Masters in Teaching Program 2006-2008: Teaching the Child in Front of You in a Changing World, The Evergreen State College, Olympia, Washington.

Purdie, N., & Hattie, J. (1996). Cultural Differences in the Use of Strategies for Self-Regulated Learning. *American Educational Research Journal, 33*(4), 845-871.

Purkey, S. C., & Smith, M. S. (1983). Effective Schools: A Review. *The Elementary School Journal, 83*(4), 426-452.

Puustinen, M., & Pulkkinen, L. (2001). Models of Self-regulated Learning:a review. *Scandinavian Journal of Educational Research, 45*(3), 269-286.

Rasch, G. (1980). *Probabilistic models for Some Intellignece and Attainment Tests*. Chicago: The University of Chicago Press.

Rea, L. M., & Parker, R. A. (2005). *Designing & Conducting Survey Research: A Comprehensive Guide* (3 ed.). CA: Jossey-Bass.

Reckase, M. D. (2002). *Contributions of Background Questions to Improving the Precision of NAEP Results*. Washington, D.C: National Assessment Governing Board.

Reichardt, C. S., & Rallis, S. F. (1994). Qualitative and quantitative inquiries are not incompatible: A call for a new partnership. In C. S. Reichardt & S. F. Rallis (Eds.), *The qualitative-quantitative debate: New perspectives* (pp. 85-92). San Francisco: Jossey-Bass.

Resnick, L. B. (1987). *Education and learning to think*. Washington, D.C: National Academy Press.

Reynolds, D. (2006). World Class Schools: Some methodical and substantive findings and implications of the International School Effectiveness Project (ISERP). *Educational Research and Evaluation, 12*(6), 535-560.

Reynolds, D., Creemers, B. P. M., Nesselrodt, P. S., Schaffer, E. C., Stringfield, S., & Teddlie, C. (1994). *Advances in School Effectiveness Research and Practice*. New York: Elsevier Science Ltd.

Reynolds, D., Teddlie, C., Creemers, B., Scheerens, J., & Townsend, T. (2000). An Introduction to School Effectiveness Research. In C. Teddlie & D. Reynolds (Eds.), *The International Handbook of School Effectiveness Research* (pp. 3). London: Falmer SeriesTaylor & Francis Group.

Ritblatt, S. N., Beatty, J. R., Cronan, T. A., & Ochoa, A. M. (2002). Relationships among Perceptions of Parent Involvement, Time Allocation, and Demographic Characteristics: Implication for Policy Formation. *Journal of Community Psychology, 30*(5), 519-549.

Ross, J. A. (1992). Teacher Efficacy and the Effects of Coaching on Student Achievement. *Canadian Journal of Education, 17*(1), 51-65.

Ross, K. N. (1992). Sample Design for International Studies of Educational Achievement. *Prospects, 22*(3), 305-316.

Ross, K. N. (2005). *Module 3: Sample Design for Educational Survey Research*. Retrieved from http://www.unesco.org/iiep.

Royal Government of Bhutan. (2005). *Bhutan National Human Development Report 2005*. Retrieved from http://www.bhutan.gov.bt/government/publications.php?av_id=0.

Ruthven, K., Hennessy, S., & Brindley, S. (2004). Teacher representations of the successful use of computer-based tools and resources in secondary-school English, mathematics and science. *Teaching and Teacher Education, 20*(3), 259-275.

Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key Characteristics of Effective Schools: A Review of School Effectiveness Research*. London: Institute of Education, University of London, England.

Scheaffer, R. L., Mendenhall(III), W., & Ott, R. L. (2006). *Elementary Survey Sampling* (6 ed.). Belmont: Wadsworth Publishing Company.

Scheerens, J. (1990). School Effectiveness Research and Development of Process Indicators of School Functioning. *School Effectiveness and School Improvement, 1*(1), 61-80.

Scheerens, J. (1992). *Effective Schooling: Research, Theory and Practice*. London: Cassell.

Scheerens, J. (1997). Conceptual Models and Theory-Embedded Principles on Effective Schooling. *School Effectiveness and School Improvement, 8*(3), 296-310.

Scheerens, J. (2000). *Improving School Effectiveness*. Retrieved from http://www.unesco.org/iiep.

Scheerens, J. (2001). Monitoring School Effectiveness in Developing Countries. *School Effectiveness and School Improvement, 12*(4), 359-384.

Scheerens, J., & Bosker, R. J. (1997). *The Foundations of Educational Effectiveness* (1st ed.). Oxford, U.K: Elsevier Science Ltd, Pergamon.

Scheerens, J., Glass, C., & Thomas, S. M. (2003). *Education Evaluation, Assessment, and Monitoring: A Systemic Approach*. Lisse: Sweets & Zeitlinger B.V., The Netherlands.

Schiller, K. S. (1999). Effects of Feeder Patterns on Students' Transition to High School. *Sociology of Education, 72*(4), 216-233.

Schleicher, A. (2000). Monitoring Student Knowledge and Skills: The OECD Programme for International Student Assessment. In D. Shorrocks-Taylor & E. Jenkin, W. (Eds.), *Learning From Others* (pp. 63-77). The Netherlands: Kluwer Academic Publishers.

Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of Multitrait-Multimethod Matrices. *Applied Psychological Measurement, 10*(1), 1-12.

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T.-Y., & Hsuan, Y.-L. (2007). A Meta-Analysis of National Research: Effects of Teaching Strategies on Student Achievement in Science in the United States. *Journal of Research in Science Teaching, 44*(10), 1436-1460.

Schunk, D. H. (1991). Self-Efficacy and Academic Motivation. *Educational Psychologist, 26*(3 & 4), 207-231.

Schunk, D. H. (1996). Goal and Self-Evaluative Influences During Children's Cognitive Skill Learning. *American Educational Research Journal, 33*(2), 359-382.

Scrimshaw, P. (2004). Enabling Teachers to Make Successful Use of ICT. Retrieved 17-10-2008, from British Educational Communications and Technology Agency (Becta): http://partners.becta.org.uk/upload-dir/downloads/page_documents/research/enablers.pdf

Secretary of Labour's Commission on Achieving Necessary Skills. (1991). *What work requires of schools: A SCANS report for America 2000*. Washington, D.C.: United States Department of Labor.

Seidel, T., & Shavelson, R. J. (2007). Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of Educational Research, 77*(4), 454-499.

Simon, B. S. (2001). Family Involvement in High School: Predictors and Effects. *NASSP Bulletin, 85*(627), 8-19.

Simon, B. S. (2004). High school outreach and family involvement. *Social Psychology of Education, 7*(2), 185-209.

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based Practices in Classroom Management: Considerations for Research to Practice. *Education and Treatment of Children, 31*(3), 351-380.

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research, 75*(3), 417-453.

Slavin, R. E. (1990). Ability Grouping and Student Achievement in Elementary Schools: A Best-Evidence Synthesis. *Review of Educational Research, 60*(3), 471-499.

Smith, J. L. (2006). The Interplay among Stereotypes, Performance-Avoidance Goals, and Women's Math Performance Expectations. *Sex Roles, 54*(3/4), 287-296.

Smith, R. (1995). Staff appraisal in higher education- a study of performance review at Nene College, Northampton. *Higher Education, 30*, 189-205.

Soderstrom, M., & Uusitalo, R. (2010). School Choice and Segregation: Evidence from an Admissionn Reform. *Scandinavian Journal of Economics, 112*(1), 55-76.

SPSS Inc. (2009). SPSS for Windows (Version 18.0). Chicago: SPSS Inc.

Stage, S. A., & Quiroz, D. R. (1997). A meta-analysis of interventions to decrease disruptive classroom behaviour in public education settings. *School Psychology Review, 26*(3), 333-368.

Stapel, D. A., & Koomen, W. (2005). Competition, Cooperation, and the Effects of Others on Me. *Journal of Personality and Social Psychology, 88*(6), 1029-1038.

Starkey, L., Yates, A., Meyer, L. H., Hall, C., Taylor, M., Stevens, S., et al. (2009). Professional Development Design: Embedding Educational Reform in New Zealand. *Teaching and Teacher Education, 25*, 181-189.

Stecher, B. M. (2002). Consequences of Large-Scale, High-Stakes Testing on School and Classroom. In L. Hamilton, et al, (Ed.), *Making sense of Test-Based Accountability in Education* (pp. 79-100): RAND Corporation.

Steele, C. M. (1997). A Threat in the Air : How Stereotypes Shape Intellectual Identity and Performance. *American Psychologist, 52*(6), 613-629.

Steenkamp, J.-B. E. M., & van-Trijp, H. C. M. (1991). The use of LISREL in validating marketing constructs. *International Journal of Research in Marketing, 8*(4), 283-299.

Sullivan, P., McDonough, A., & Prain, V. (n.d). Student Engagement in the Middle Years: Describing Influences and Possible Teacher Actions. 1-8. Retrieved from http://www.aare.edu.au/05pap/sul05134.pdf

Tashakkori, A., & Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Thousand Oaks: Sage Publications, Inc.

Taylor, S. D. (2000). International Comparisons of Pupil Performance *An Introduction and Discussion*. In D. Taylor, Shorrocks, & E. Janovy, W. (Eds.), *Learning from Others* (pp. 13-27). The Netherlands: Kluwer Academic Publishers.

Taylor, S. D., Jenkin, E. W., Curry, J., & Swennerton, B. (2000). International Comparisons of Pupil Performance and national Mathematics and Sceince Testing in England. In K. Tobin (Ed.), *Learning from Others* (pp. 233-258). The Netherlands: Kluwer Academic Publishers.

Teddlie, C., & Reynolds, D. (Eds.). (2000). *The International Handbook of School Effectiveness Research*. London: Palmer Press.

Thomson, S., & Bortoli, L. D. (2008). Exploring scientific literacy; how Australia measures up: the PISA 2006 survey of students' scientific, reading and mathematical literacy skills. In W. Whitham (Eds.), PISA National reports Available from http://www.acer.edu.au/ozpisa/reports.html

Thrupp, M., & Lupton, R. (2006). Taking School Contexts more Seriously: The Social Justice Challenge. *British Journal of Educational Studies, 54*(3).

Tileston, D. W. (2010). *What Every Teacher Schould Know About Student Motivation* (2 ed.). London: SAGE Ltd.

Timar, T. B., & Roza, M. (2010). "A False Dilema": Schould Decisions about Education Resource Use Be Made at the State or Local Level. *American Journal of Education, 116*(3), 397-422.

Tschannen-Moran, M., Hoy, A. W., & Hoay, W. K. (1998). Teacher Efficacy: Its Meaning and Measures. *Review of Educational Research, 68*(2), 202-248.

Tshering, G. (2006). *IRT in Item Banking, Study of DIF Items and Test Construction.* Unpublished Master's Thesis, University of Twente, The Netherlands.

Tshering, G., & Prain, V. (2011). Benchmarking the performance of Bhutanese students with the performance of the students from the OECD's PISA countries. *Educational Research and Evaluation, 17*(4), 263-281.

Tuijnman, C. A., Kirsch, I. S., & Wagner, D. A. (Eds.). (1997). *Adult basic skills: Advances in measurement and policy*. Creskill, New York: Hampton.

UNESCO. (1997). ISCED 1997, International Standard Classification of Education Available from www.uis.unesco.org/TEMPLATE/pdf/isced/ISCED_A.pdf

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The Relation Between Self-Beliefs and Academic Achievement: A Meta-Analytic Review. *Educational Psychologist, 2*(39), 111-133.

Vansteenkiste, M., Timmermans, T., Lens, W., Soenens, B., & Broeck, A. V. d. (2008). Does Extrinsic Goal Framing Enhance Extrinsic Goal-Oriented Individuals’ Learning and Performance? An Experimental Test of the Match Perspective Versus Self-Determination Theory *Journal of Educational Psychology, 100*(2), 387-397.

Wagemaker, H. (2008). Choices and trade-offs: reply to McGaw. *Assessment in Education: Principles, Policy & Practice, 15*(3), 267-278.

Walberg, H. J. (1980). A Psychological Theory of Educational Productivity. In F. H. Farley & N. Gordon (Eds.), *Psychology and Education* (pp. 81-110). Chicago: National Society for the Study of Education.

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership, 41*(8), 19-27.

Walberg, H. J. (2002). *The National Assessment's Most Useful Background Items*. Washington, D.C: National Assessment Governing Board.

Walberg, H. J., Haertel, G. D., Pascarella, E., Junker, L. K., & Boulanger, F. D. (1981). Probing a Model of Educational Productivity in Science with national Assessment Samples of Early Adolescents. *American Educational Research Journal, 18*(2), 233-249.

Waslander, S., Pater, C., & Weide, M. v. d. (2010). *Markets in Education: An Analytical Review of Empirical Research on Market Mechanisms in Education* (No. Education Working Papers, No 52). Paris: OECD.

Watanabe, R. (1992). How Japan makes use of international education survey research. *Prospects, 22*(4), 456-462.

Wayne, A. J., Yoon, K. S., Zhu, P., Cronon, S., & Garet, M. S. (2008). Experimenting With Teacher Professional Development: Motives and Methods. *Educational Researcher, 37*(8), 469-479.

Wayne, A. J., & Youngs, P. (2003). Teacher Characteristics and Student Achievement Gains: A Review. *Review of Educational Research, 73*(1), 89-122.

Weaver-Hightower, M. (2003). The "Boy Turn" in Research on Gender and Education. *Review of Educational Research, 73*(4), 471-498.

Webb, N. L. (1997). *Research Monograph No.8: Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education*. Washington, DC: Council of Chief State School Officers.

Webb, N. L. (1999). *Research Monograph No. 18: Alignment of Science and Mathematics Standards and Assessments in Four States*. Madison, WI: National Institute for Science Education.

Webb, N. L. (2006). Identifying Content for Student Achievement Tests. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 155-180). Mahwah, New Jersey: Lawrence Erlbaum, Inc.

Weller, L. D. (1996). Benchmarking: a paradigm for change to quality education. *The TQM Magazine*, *8*(6), 24-29.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Educational Policy Analysis Archives, 10(12).* Retrieved 2 June 2008, from http://epaa.asu.edu/epaa/v10n12/

West, A., Allmendinger, J., Nikolai, R., & Barham, E. (2010). Decentralization and educational achievement in Germany and the UK. *Environment and Planning C: Government and Policy, 28*, 450-468.

White, K. R. (1982). The Relation Between Socioeconomic Status and Academic Achievement. *Psychological Bulletin, 91*(3), 461-481.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-Value Theory of Achievement Motivation. *Contemporary Educational Psychology, 25*, 68-81.

Wiggan, G. (2007). Race, School Achievement, and Educational Inequality: Toward a Student-Based Inquiry Perspective. *Review of Educational Research, 77*(3), 310-333.

Wiley, D. E., & Wolfe, R. G. (1992). Major survey design issues for the IEA Third International Mathematics and Science Study. *Prospects, 22*(3), 297-304.

Wiliam, D. (2008). International comparisons and sentivity to instruction. *Assessment in Education: Principles, Policy & Practice, 15*(3), 253-257.

Wilson, D. (2004). The Interface of School Climate and School Connectedness and Relationships with Aggression and Victimization. *The Journal of School Health, 74*(7), 293-299.

Wise, K., & Okey, J. R. (1983). A Meta-Analysis of the Effects of various Science Teaching Strategies on Achievement. *Journal of Research in Science Teaching*, *20*(5), 419-435.

Witziers, B., Bosker, R. J., & Kruger, M. L. (2003). Educational Leadership and Student Achievement: The Elusive search for an Association. *Educational Administration Quarterly, 39*(3), 398-425.

Wobmann, L. (2007). International Evidence on School Competition, Autonomy, and Accountability: A Review. *Peabody Journal of Education, 82*(2), 473-497.

Wong, K.-c. (2005). Conditions and practices of successful principalship in Shanghai. *Journal of Educational Administration, 43*(6), 552-562.

Wu, M. L. (2005). The Role of Plausible Values in Large-Scale Surveys. *Studies in Educational Evaluation, 31*(2-3), 114-128.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest: Generalised Item Response Modelling Software* (Vol. 2). Melbourne: ACER Press.

Yan, W., & Lin, Q. (2005). Parent Involvement and mathematics Achievement: Contrast Across Racial and Ethnic Groups. *The Journal of Educational Research, 99*(2), 116-127.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (No. 033). Washington, DC: Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Yoon, K. S., Garet, M., Birman, B., & Jacobson, R. (2006). *Examining the Effects of Mathematics and Science Professional Development on Teachers' Instructional Practice: Using Professional Development Activity Log.* Washington, DC: Council of Chief State School Officers.

Younger, M. R., & Warrington, M. (2006). Would Harry and Hermione Have Done Better in Single-Sex Classes? A Review of Single-Sex Teaching in Coeducational Secondary Schools in the United Kingdom. *American Educational Research Journal*, *43*(2), 579-620.

Younger, M. R., & Warrington, M. (2007). Closing the Gender Gap? Issues of gender equity in English secondary schools. *Discourse: Studies in the cultural politics of education*, *28*(2), 219-242.

Zimmerman, B. J. (1989). A social Cognitive View of Self-Regulated Academic Learning. *Journal of Educational Psychology, 81*(3), 329-339.

Zimmerman, B. J. (1990). Self-Regulated Learning and Academic Achievement: An Overview. *Educational Psychologist*, *25*(1), 3-17.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-Motivation for Academic Attainment: The Role of Self-Efficacy Beliefs and Personal Goal Setting. *American Educational Research Journal*, *29*(3), 663-676.

Zimmerman, B. J., & Martinez-Pons, M. (1988). Construct Validation of a Strategy model of Student Self-Regulated Learning. *Journal of Educational Psychology, 80*(3), 284-290.

Zuckerman, J. T. (2007). Classroom Management in Secondary Schools: A Study of Student Teachers' Successful Strategies. *American Secondary Education, 35*(2), 4-16.